

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:05 ; Search time 2378.87 Seconds  
(without alignments)  
633.372 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_107  
Perfect score: 72  
Sequence: 1 atggcctctctggcctcca.....ttttgggcacactggttgc 72

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	72	100.0	615	6	AX136475
2	72	100.0	693	9	AF250558
3	72	100.0	1441	6	AX286822
4	72	100.0	1475	6	AX092348
5	72	100.0	1475	6	AX299996
6	72	100.0	1506	9	BC014424
7	72	100.0	1918	9	AF177340
8	72	100.0	2863	6	AX136129
9	72	100.0	109465	9	AL158821
10	62.4	86.7	953	4	AF358907
11	56	77.8	49323	2	AC107455
12	54.4	75.6	779	6	AX286824
13	54.4	75.6	791	6	E31591
14	54.4	75.6	791	10	AF072128
15	54.4	75.6	2828	10	BC015252
16	39.4	54.7	51	6	AX165471
17	29.6	41.1	198656	2	AC091323
18	29.4	40.8	51	6	AX165593
19	28.8	40.0	720	9	HSAL32445
20	28.8	40.0	1233	9	BC012126
21	28.8	40.0	1938	9	AF314090
22	28.8	40.0	71409	5	AC091535
23	28.8	40.0	75015	5	AC091626
24	28.8	40.0	100058	9	AP000695
25	28.8	40.0	101608	9	AP000694
26	28.8	40.0	120000	2	HSAC000005
27	28.8	40.0	175835	2	AC024074
28	28.8	40.0	340000	9	AP001726
29	28.6	39.7	1851	9	AF408152
30	28.6	39.7	1859	9	AF408151
31	28.6	39.7	1905	9	AF408153
32	28.4	39.4	3643	10	MUSAP1
33	28.4	39.4	36901	6	AX015903
34	28.4	39.4	38886	6	AX015909
35	28.4	39.4	228283	10	AF312994
36	28	38.9	4046	4	DGACADENCYC
37	28	38.9	4046	6	I29958
38	28	38.9	173762	2	AC012572
39	28	38.9	199254	2	AC021144
40	28	38.9	202032	2	AC087399
41	27.8	38.6	728	4	ECU60978
42	27.6	38.3	90109	2	AC002557
43	27.4	38.1	157041	2	AP004363
44	27.4	38.1	165110	9	AL450311
45	27.4	38.1	193219	10	AL606975

ALIGNMENTS

RESULT	1
LOCUS	AX136475
DEFINITION	Sequence 397 from Patent EP1067182.
ACCESSION	AX136475
VERSION	AX136475.1
KEYWORDS	GI:14272879
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 615)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.

TITLE Secretory protein or membrane protein  
JOURNAL Patent: Ep 1067182-A 397 10-JAN-2001;  
Helix Research Institute (JP)

FEATURES Location/Qualifiers  
source 1..615  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 133 a 174 c 157 g 143 t 8 others  
ORIGIN -----

Query Match 100.0%; Score 72; DB 6; Length 615;  
Best Local Similarity 100.0%; Pred. No. 3.9e-14;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttggtgacactgttggtgacatcctagcctcttggtgac 60  
|||||  
Db 231 ATGGCCTCTCTTGCGCTCCACCTGTGGGCTACATCTAGCGCTTCTGCGGCTTTGGGC 290  
|||||

QY 61 acactggtgcc 72  
|||||  
Db 291 ACACCTGCTTGCC 302  
|||||

RESULT 2  
AF250558 693 bp mRNA linear PRI 09-AUG-2000  
LOCUS Homo sapiens claudin-2 mRNA, complete cds.  
DEFINITION AF250558  
ACCESSION AF250558  
VERSION AF250558.1 GI:9755008  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
AUTHORS Reinecker,H.-C., Sakaguchi,T. and Golden,H.M.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2000) Gastrointestinal Unit, Massachusetts  
General Hospital, Fruit Street, Boston, MA 02114, USA

FEATURES  
source Location/Qualifiers  
1..693  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq22.3-q23"  
/cell\_line="T-84"  
/tissue\_type="epithelium"  
1..693  
/note="tight junction-associated protein"  
/codon\_start=1  
/product="claudin-2"  
/protein\_id="AA98151.1"  
/db\_xref="GI:9755009"  
/translation="MASIGQLVGYIIGLGLTFLVAMLPSPKTSYVGSATYAV  
GFSKGLMMECATSTGTGCDIYFTILGLPLDPAQAAOMYTSASISLACIIVGM  
RCTVCEOSRAKRDVAVAGVFTIIGLGLIPVAMNHGILRDYSPVPSMKFEI  
GEALYLGIISSFLIAGIILICFSCSSORNSNTIDAYQADPLATRSSRPGQPPVK  
SEFNSYSLTGIV"

BASE COUNT 142 a 202 c 168 g 181 t

ORIGIN

Query Match 100.0%; Score 72; DB 9; Length 693;  
Best Local Similarity 100.0%; Pred. No. 3.9e-14;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttggtgacactgttggtgacatcctagcctcttggtgac 60  
|||||  
Db 1 ATGGCCTCTCTTGCGCTCCACCTGTGGGCTACATCTAGCGCTTCTGCGGCTTTGGGC 60  
|||||

QY 61 acactggtgcc 72  
|||||  
Db 61 ACACCTGCTTGCC 72  
|||||

RESULT 3  
AX286822 1441 bp DNA linear PAT 21-NOV-2001  
LOCUS AX286822  
DEFINITION Sequence 1 from Patent WO0180879.  
ACCESSION AX286822  
VERSION AX286822.1 GI:17048854

KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Vinals y de Bassols,C.G., Gaulis,S.R., Coche,T.G. and Orntoft,T.A.  
TITLE Colorectal cancer vaccines and diagnosis  
JOURNAL Patent: WO 0180879-A 1 01-NOV-2001;  
SMITHKLINE BEECHAM BIOLOGICALS (S.A.)

FEATURES  
source Location/Qualifiers  
1..1441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 337 a 399 c 357 g 345 t 3 others

ORIGIN

Query Match 100.0%; Score 72; DB 6; Length 1441;  
Best Local Similarity 100.0%; Pred. No. 4.3e-14;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttggtgacactgttggtgacatcctagcctcttggtgac 60  
|||||  
Db 77 ATGGCCTCTCTTGCGCTCCACCTGTGGGCTACATCTAGCGCTTCTGCGGCTTTGGGC 136  
|||||

QY 61 acactggtgcc 72  
|||||  
Db 137 ACACCTGCTTGCC 148  
|||||

RESULT 4  
AX092348 1475 bp DNA linear PAT 21-MAR-2001  
LOCUS AX092348  
DEFINITION Sequence 79 from Patent WO0116318.  
ACCESSION AX092348  
VERSION AX092348.1 GI:13444488  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1475)  
AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,  
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
Wood,W.I.  
TITLE Secretd and transmembrane polypeptides and nucleic acids encoding  
JOURNAL the same  
Patent: WO 0116318-A 79 08-MAR-2001;  
Genentech, Inc. (US)

FEATURES  
source Location/Qualifiers  
1..1475  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 345 a 404 c 376 g 350 t

ORIGIN

Query Match 100.0%; Score 72; DB 6; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 4.3e-14;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttggtgacactgttggtgacatcctagcctcttggtgac 60  
|||||  
Db 122 ATGGCCTCTCTTGCGCTCCACCTGTGGGCTACATCTAGCGCTTCTGCGGCTTTGGGC 181  
|||||

QY 61 acactggtgcc 72  
|||||  
Db 182 ACACCTGCTTGCC 193  
|||||

RESULT 5  
AX299996

LOCUS AX299996 1475 bp DNA linear PAT 26-NOV-2001  
 DEFINITION Sequence 1 from Patent WO0166740.  
 ACCESSION AX299996  
 VERSION AX299996.1 GI:17129473  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Eaton,D.L., Fong,S., Goddard,A., Godowski,P.J., Grimaldi,C.J.,  
 Guney,A.L., Tumas,D., Watanabe,C.K., Wood,W.I. and Zhang,Z.  
 TITLE Compositions and methods for the treatment of immune related  
 diseases  
 JOURNAL Patent: WO 0166740-A 1 13-SEP-2001;  
 Genentech, Inc. (US)  
 FEATURES  
 Source  
 1..1475  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 345 a 404 c 376 g 350 t  
 ORIGIN  
 Query Match 100.0%; Score 72; DB 6; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-14;  
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atggcctcttggcctccaaactgtggctacatctagcctctgggctttgggc 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 122 ATGGCTCTCTGGCCCTCCAACTTGCGGTACATCTAGGCCCTCTGGGCTTTGGGC 181  
 QY 61 acactggttgcc 72  
 ||||||||||||  
 Db 182 ACACGTGGTGCC 193  
 RESULT 6  
 BC014424  
 LOCUS BC014424 1506 bp mRNA linear PRI 19-SEP-2001  
 DEFINITION Homo sapiens, similar to claudin 2, clone MGC:20191 IMAGE:4645075,  
 mRNA, complete cds.  
 ACCESSION BC014424  
 VERSION BC014424.1 GI:15680158  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1506)  
 AUTHORS Strausberg,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-SEP-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ARCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcysc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline  
 Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,  
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 29 Row: m Column: 11.  
 Location/Qualifiers

## FEATURES

## source

1..1506  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:20191 IMAGE:4645075"  
 /tissue\_type="Colon, adenocarcinoma"  
 /clone\_lib="NIH\_MGC\_15"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 122..814  
 /codon\_start=1  
 /product="Similar to claudin 2"  
 /protein\_id="AAH14424.1"  
 /db\_xref="GI:15680158"  
 /translation="MASGLQVLGVYILGLLGLTAVMLPSWKTSSYVGASIVTAV  
 GFSGLWMEKATHSTGITQCDIYTLGLPADIQAAQAMMTSSAISLACIISVVG  
 RCTVFCQESRAKDRVAVAGGVFFILGLLGFIPVAVNLHGLRDFYSPLPDMSKFEI  
 GEALYLGIISSLSLIAGIILCFSCSSORNESNYDAYQAQPLATRSRPGQPKVK  
 SEFNYSILTVV"

## CDS

BASE COUNT 377 a 404 c 376 g 349 t  
 ORIGIN

Query Match 100.0%; Score 72; DB 9; Length 1506;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-14;  
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atggcctcttggcctccaaactgtggctacatctagcctctgggctttgggc 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 122 ATGGCTCTCTGGGCTCCAACTTGCGGTACATCTAGGCCCTCTGGGCTTTGGGC 181  
 QY 61 acactggttgcc 72  
 ||||||||||||  
 Db 182 ACACGTGGTGCC 193

RESULT 7  
 AF177340  
 LOCUS AF177340 1918 bp mRNA linear PRI 03-OCT-2000  
 DEFINITION Homo sapiens clone SP82 claudin 2 mRNA, complete cds.  
 ACCESSION AF177340  
 VERSION AF177340.1 GI:10503979

KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1918)  
 AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
 Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
 Yu,J. and Han,L.H.  
 TITLE Novel human cDNA clone with function of inhibiting cancer cell  
 growth

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1918)  
 AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
 Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
 Yu,J. and Han,L.H.  
 TITLE Direct Submission

JOURNAL Submitted (06-AUG-1999) National Laboratory For Oncogenes & Related  
 Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie Tu Road, Shanghai  
 200032, P.R. China  
 Location/Qualifiers

## FEATURES

## source

1..1918  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="SP82"  
 520..1212  
 /note="similar to Mus musculus claudin 2"

## CDS



[illegible]

LOCUS AF358907 953 bp mRNA linear MAR 08-MAY-2001  
 DEFINITION Canis familiaris integral membrane protein claudin-2 mRNA, complete cds.  
 ACCESSION AF358907  
 VERSION AF358907.1 GI:13991612  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 REFERENCE  
 AUTHORS Furuse, M., Furuse, K., Sasaki, H. and Tsukita, S.  
 TITLE 1 (bases 1 to 953)  
 JOURNAL Conversion of zonulae occludentes from tight to leaky strand type by introducing claudin-2 into Madin-Darby canine kidney 1 cells  
 JOURNAL J. Cell Biol. 153 (2), 263-272 (2001)  
 MEDLINE 21206012  
 PUBMED 11309408  
 REFERENCE  
 2 (bases 1 to 953)  
 AUTHORS Furuse, M., Furuse, K., Sasaki, H. and Tsukita, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAR-2001) Cell Biology, Kyoto University, Faculty of Medicine, Yoshida-Kono, Sakyo, Kyoto 606-8501, Japan  
 FEATURES  
 source  
 1. 953  
 /organism="Canis familiaris"  
 /db\_xref="taxon:9615"  
 /sex="male"  
 /tissue\_type="liver"  
 59. 751  
 /note="localizes to tight junctions"  
 /codon\_start=1  
 /product="integral membrane protein claudin-2"  
 /protein\_id="AAK51433.1"  
 /db\_xref="GI:13991613"  
 /translation="MASIGLGVYIIIGLIGLVAMLPSPMRSSVVGSIYAV  
 GFSGLMNECHTSTGTQCDIYSLILPADIQAAQAMATYSSAISLACTSVYGM  
 RCTVCFQDSRAKDLAVGVGFITIGLGLFIPVAMNLHGIIRDSYPLPQSPKFEI  
 GEALYLGIISSLSFISVAGIILCFSCPIQGNRSDYDSYOAPLATRGSPPQPPRAK  
 SEFNSYSTATGYV"  
 BASE COUNT 195 a 289 c 265 g 204 t  
 ORIGIN  
 Query Match 86.7%; Score 62.4; DB 4; Length 953;  
 Best Local Similarity 91.7%; Pred. No. 7.5e-11;  
 Matches 66; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 atggcctctctggtcccaactgtggctacatcctagccttgggctttgggc 60  
 DB 59 atggcctctctggtcccaactgtggctacatcctagccttgggctttgggc 118  
 QY 61 acaactggttgc 72  
 DB 119 accctgtgtgcc 130  
 RESULT 11  
 AC107455 : 49323 bp DNA linear HTG 21-JAN-2002  
 LOCUS Mus musculus clone RP24-68D11, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC107455  
 ACCESSION AC107455  
 VERSION AC107455.1 GI:18252733  
 KEYWORDS HTG; HTGS; PHASE0.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 49323)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Unpublished  
 JOURNAL Mus musculus, clone RP24-68D11  
 REFERENCE  
 2 (bases 1 to 49323)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,

TITLE JOURNAL  
 COMMENT  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collum, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamel, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nordu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Vael, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (21-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A. F. A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L19885  
 Center clone name: 68\_D\_11  
 -----  
 \* NOTE: This record contains 62 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1  
 \* 703 802: contig of 702 bp in length  
 \* 703 802: gap of 100 bp  
 \* 803 1513: contig of 711 bp in length  
 \* 1514 1613: gap of 100 bp  
 \* 1614 2279: contig of 666 bp in length  
 \* 2280 2379: gap of 100 bp  
 \* 2380 3087: contig of 708 bp in length  
 \* 3088 3187: gap of 100 bp  
 \* 3188 3885: contig of 699 bp in length  
 \* 3887 3986: gap of 100 bp  
 \* 3987 4687: contig of 701 bp in length  
 \* 4688 4787: gap of 100 bp  
 \* 4788 5473: contig of 686 bp in length  
 \* 5474 5573: gap of 100 bp  
 \* 5574 6271: contig of 698 bp in length  
 \* 6272 6371: gap of 100 bp  
 \* 6372 7070: contig of 699 bp in length  
 \* 7071 7170: gap of 100 bp  
 \* 7171 7868: contig of 686 bp in length  
 \* 7869 7968: gap of 100 bp  
 \* 7969 8665: contig of 697 bp in length  
 \* 8666 8765: gap of 100 bp  
 \* 8766 9473: contig of 708 bp in length  
 \* 9474 9573: gap of 100 bp  
 \* 9574 10264: contig of 691 bp in length  
 \* 10265 10364: gap of 100 bp



Matches	61.	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
Qy	1	atggcctctcttggcctccaaactgtggtgcctacacatcctagcctctcttggggctttgggc	60						
Db	48	ATGGCCTCCCTTTGGGGTTCACCTGGTGGCTACATCTTCATGAGCCCTTTTGGGGCTGTAGGC	107						
Qy	61	acactggtgcgc	72						
Db	108	ACATCCATTGCC	119						
RESULT	13								
LOCUS	E31591	791 bp	DNA	Linear	PAT 07-FEB-2001				
DEFINITION	Tight junction-constituting membrane protein claudin family.								
ACCESSION	E31591								
VERSION	E31591.1	GI:13018519							
KEYWORDS	JP 2000032984-A/2.								
SDUCE	Mus sp.								
ORGANISM	Mus sp.								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
AUTHORS	1 (bases 1 to 791)								
TITLE	Syolichiro.T.								
JOURNAL	Tight junction-constituting membrane protein claudin family								
COMMENT	Patent: JP 2000032984-A 2 02-FEB-2000; EISAI CO LTD OS Mus sp. (mouse) PN JP 2000032984-A/2 PD 02-FEB-2000 PF 26-JUN-1998 JP 1998179847 PR								
FEATURES	PI SYOICHIRO TSUKITA PC .C12N15/09.C07K14/47.C07R16/18.C12N5/10.C12P21/02.C12Q1/02. PC C12Q1/68. PC .G01N33/15.G01N33/53.G01N33/577//C12P21/08.(C12N5/10.C12R1.91). PC (C12P21/02.C12R1.91).C12N15/00.C12N5/00.(C12N5/00.C12R1.91). CC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers FT CDS 1..791 /organism="Mus sp." /db_xref="taxon:10095"								
BASE COUNT	155 a 207 c 220 g 209 t								
ORIGIN									
Query Match	75.6%.	Score 54.4;	DB 6;	Length 791;					
Best Local Similarity	84.7%.	Pred. No. 3.8e-08;							
Matches	61; Conservative	0; Mismatches	11; Indels	0; Gaps	0;				
Qy	1	atggcctctcttggcctccaaactgtggtgcctacacatcctagcctctcttggggctttgggc	60						
Db	48	ATGGCCTCCCTTTGGGGTTCACCTGGTGGCTACATCTTCATGAGCCCTTTTGGGGCTGTAGGC	107						
Qy	61	acactggtgcgc	72						
Db	108	ACATCCATTGCC	119						
RESULT	14								
LOCUS	AF072128	791 bp	mRNA	Linear	ROD 21-JUL-1998				
DEFINITION	Mus musculus claudin-2 mRNA, complete cds.								
ACCESSION	AF072128								
VERSION	AF072128.1	GI:3335183							
KEYWORDS									
SDUCE	house mouse.								
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								

FEATURES	source	CDs
REFERENCE AUTHORS	1 (bases 1 to 791) Furuse,M., Fujita,K., Hiragaki,T., Fujimoto,K. and Tsukita,S.	
TITLE	Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin	
JOURNAL	J. Cell Biol. 141 (7), 1539-1550 (1998)	
MEDLINE	98311639	
REFERENCE AUTHORS	2 (bases 1 to 791) Furuse,M., Fujita,K., Hiragaki,T., Fujimoto,K. and Tsukita,S.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-JUN-1998) Cell Biology, Kyoto University Faculty of Medicine, Sakyo-Ku, Kyoto 606-8501, Japan	
FEATURES	location/Qualifiers	
source	1..791	
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	48..740	
	/note="Integral membrane protein localizing at tight junctions"	
	/codon_start=1	
	/product="Claudin-2"	
	/protein_id="AAC27079.1"	
	/db_xref="GI:3335184"	
	/translation="MASIGVQLVGVILGLLGLTSTAMLLPNRRTSSYVGASIVTAVGFSGLMMECATNHTGTQCDIVSTLGLPADIQAOAMVNTSSAMSILACIISVGMCEATYFCODSRADKDVAVVGGVFFLIGLIPANMLHGIIRDFVPIVPSDKPERI	
	GCALYGLIISALPSLVAGVILCFSCSPQGNRTNYIDYIQAOPLATRSSPNSAQQPKAKSEFNSISLTGIV"	
BASE COUNT	155 a 207 c 220 g 209 t	
ORIGIN		
Query Match	75.6%; Score 54.4; DB 10; Length 791;	
Best Local Similarity	84.7%; Pident. No.3.8e-08;	
Matches	61; Conservative 0; Mismatches 11; Gaps 0;	
Oy	1 atggcctctcttgagcctcaactcttgaggagccacatcctagagcctcttgaggcttttggc 60	
Db	48 ATGGCCTCCCTTGGCGCTTCACACTGCTGAGGCTACATCTTAGGCCCTTTGGGCTTTAGGC 107	
Oy	61 aaactggtgcc 72	
Db	108 AATTCATTTGCC 119	
RESULT 15		
LOCUS	BC015252	
DEFINITION	Mus musculus, claudin 2, clone MGC:18348 IMAGE:4223446, mRNA,	
ACCESSION	BC015252	
VERSION	BC015252.1 GI:15929633	
KEYWORDS	house mouse.	
SDUCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE AUTHORS	1 (bases 1 to 2828)	
TITLE	Strausberg,R.	
JOURNAL	Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT	Contact: MGC help desk	
	Email: cgabbs-remail.nih.gov	
	Tissue Procurement: Jeffrey E. Green, M.D.	
	cDNA Library Preparation: Life Technologies, Inc.	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNT)	
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305	
	Web site: http://www.shgc.stanford.edu	
	Contact: (Dickson, Mark) mcd@paxil.stanford.edu	

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 25 Row: h Column: 15  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7710003.

# FEATURES

source

Location/Qualifiers

1..2828

/organism="Mus musculus"

/db\_xref="LocusID:12738"

/db\_xref="taxon:10090"

/clone="MGC:18348 IMAGE:4223446"

/tissue\_type="Kidney, normal. 5 month old male mouse."

/clone\_lib="NCI\_CGAP\_Kid14"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

126..818

/codon\_start=1

/product="claudin 2"

/protein\_id="AAH15252.1"

/db\_xref="GI:15929634"

/translation="MASLGVQLVGYILGLGLGTSIAMLPPNRTSSYVGASIVTAV

GFSGLMWECATHSGITQCDIVSTLLGLPADIOAAQAMMTSSAMSSLACIISVVGGM

RCTVFCDSRAKDRVAVGVGVFLLGLGILGFIPIVANNHGLILRDEYSPLVPDSMKFEI

GEALYGLISALFSLVAGVILCFSCSPQGNRRNIYDIOAQPLATRSSPRSAQOPKAK

SEFNYSYSLTGYV"

BASE COUNT 736 a 697 c 667 g 728 t

ORIGIN

Query Match

75.6%; Score 54.4; DB 10; Length 2828;

Best Local Similarity 84.7%; Pred. No. 4.4e-08;

Matches 61; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 atggcctcttggcctcaactgtgggtacatccttagcctcttgggcttttgggc 60

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 126 ATGGCCTCCCTTGGCGTCAACTGGTGGGTACATCCTAGGCCCTTTGGGCTGTAGGC 185

Qy 61 acactggttgcc 72

||| |||||

Db 186 ACATCCATTGCC 197

Search completed: June 2, 2002, 18:58:13

Job time: 11113 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:09:02 ; Search time 523.16 Seconds  
(without alignments)  
236.291 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_107

Perfect score: 72

Sequence: 1 atggcctctctggcctcca.....ttttgggcacactggttgc 72

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72	100.0	324	20	AAAX41060
2	72	100.0	467	21	AAC98817
3	72	100.0	530	22	AAD08518
4	72	100.0	615	22	AAF93963
5	72	100.0	1400	20	AAAX97865
6	72	100.0	1441	24	AAS16180
7	72	100.0	1475	21	AAAX37060
8	72	100.0	1475	22	AAS15360
9	72	100.0	1475	22	AAS21489
					Human secreted pro
					Human pancreatic c
					Human secreted pro
					Primer specific fo
					Human secreted pro
					Human CASB81 polyn
					Human PRO1356 (UNQ
					CDNA encoding huma
					Human cDNA sequenc

10	72	100.0	1475	22	AAF92097	Human PRO1356 cDNA
11	72	100.0	1475	22	AAF54296	DNA encoding prote
12	72	100.0	1481	22	AAD08497	Human secreted pro
13	72	100.0	1524	20	AAAX97818	Human secreted pro
14	72	100.0	1540	22	AAI58013	Human polynucleoti
15	72	100.0	1761	22	AAI59799	Human polynucleoti
16	72	100.0	1918	24	ABA04424	Human SP82 protein
17	72	100.0	2742	21	AAAL12585	CDNA encoding a me
18	72	100.0	2863	22	AAF93769	Human cDNA encodin
19	71.6	99.4	405	20	AAAX1059	Human secreted pro
20	54.4	75.6	779	24	AAS16181	Murine CASB81 poly
21	54.4	75.6	791	21	AAZ89137	Murine clodin 2 cD
22	39.4	54.7	51	23	ABL00675	Human silent nonco
23	29.4	40.8	51	23	ABL00797	Human amino acid c
24	28.8	40.0	1174	21	AAA37113	Human PRO1571 (UNQ
25	28.8	40.0	1174	22	AA546101	Human DNA encoding
26	28.8	40.0	1174	22	AAF54431	Probe #47 used in
27	28.4	39.4	36901	20	AAZ23892	Murine LOBO genom
28	28.4	39.4	38886	20	AAZ23897	Murine LOBO genom
29	28	38.9	4046	14	AAQ42525	Cardiac adenylyl c
30	27.2	37.8	660	21	AAZ89153	Murine clodin 6 cD
31	27.2	37.8	1231	22	AAF32670	Human cDNA encodin
32	26.6	36.9	806	22	AA351139	Human cardiiovascul
33	26.6	36.9	818	21	AAC77281	Human ORPX ORF2836
34	26.4	36.7	1497	21	AAC40181	Arabidopsis thalia
35	26.4	36.7	1515	21	AAC50415	Arabidopsis thalia
36	26.2	36.4	1530	23	ABL07695	Drosophila melanog
37	26.2	36.4	3609	23	ABL07694	Drosophila melanog
38	26.2	36.4	6968	23	ABL21538	Drosophila melanog
39	26	36.1	1653	22	AAD21436	ADAM-20dis-Fc fusi
40	26	36.1	1981	21	AAZ50482	Corn sulphate perm
41	26	36.1	2181	20	AAZ56461	Human SVPH1-26 pro
42	26	36.1	2426	20	AAZ28636	Nucleotide sequenc
43	25.8	35.8	766	23	AA592167	DNA encoding novel
44	25.8	35.8	783	22	AAH99770	Human protein enco
45	25.8	35.8	783	22	AAI61349	Human polynucleoti

ALIGNMENTS

RESULT 1

AAAX41060  
ID AAAX41060 standard; cDNA; 324 BP.

XX  
AC AAAX41060;

XX  
DT 18-JUN-1999 (first entry)

XX  
DE Human secreted protein 5' EST SEQ ID NO: 272.

XX  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

XX  
PN WO9906554-A2.

XX  
PD 11-FEB-1999.

XX  
PF 31-JUL-1998; 98WO-IB01238.

XX  
PR 01-AUG-1997; 97US-0905134.

XX  
(GEST ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX  
DR WPI; 1999-153784/13.

DR P-PSDB; AAV12227.  
XX  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT CDNA libraries prepared from kidney, fetal kidney, dystrophic  
PT muscle, muscle and heart tissue  
XX  
PS Claim 1; Page 402-403; 622pp; English.  
XX  
CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAY01607 and  
CC AAY11994 to AAY12260, respectively. The proteins given represent the  
CC signal peptide and an N-terminal fragment of a secreted protein. The  
CC nucleic acid sequences can be used for producing secreted human gene  
CC products. They can also be used to develop products for diagnosis and  
CC therapy. The proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used  
CC for directing extracellular secretion of a polypeptide or the insertion  
CC of a polypeptide into a membrane, or importing a polypeptide into  
CC a cell.  
XX  
SQ Sequence 324 BP; 73 A; 80 C; 94 G; 74 T; 3 other;

Query Match 100.0%; Score 72; DB 20; Length 324;  
Best Local Similarity 100.0%; Pred. No. 2e-15;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctctcttggtcccaactgtggtctacatcctagcctcttggtggctttgggc 60  
|||  
Db 160 atggcctctcttggtcccaactgtggtctacatcctagcctcttggtggctttgggc 219  
|||  
OY 61 acactggttggc 72  
|||||  
Db 220 acactggttggc 231

RESULT 2  
AAC98817  
ID AAC98817 standard; cDNA; 467 BP.

AC AAC98817;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:45.  
XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antinflammatory; cardiac; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative; ss.

OS Homo sapiens.  
XX  
PN WO200055320-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05989.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-579444/54.  
DR P-PSDB; AAB54052.

PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
PS Claim 1; Page 527-528; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiac and anti-inflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides. Including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX

SQ Sequence 467 BP; 95 A; 132 C; 126 G; 109 T; 5 other;  
Query Match 100.0%; Score 72; DB 21; Length 467;  
Best Local Similarity 100.0%; Pred. No. 2.2e-15;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctctcttggtcccaactgtggtctacatcctagcctcttggtggctttgggc 60  
|||  
Db 77 atggcctctcttggtcccaactgtggtctacatcctagcctcttggtggctttgggc 136  
|||

OY 61 acactggttggc 72  
|||||  
Db 137 acactggttggc 148

RESULT 3

AAD08518  
ID AAD08518 standard; cDNA; 530 BP.

AC AAD08518;

DT 09-AUG-2001 (first entry)

DE Human secreted protein-encoding gene 10 cDNA clone HYP183, SEQ ID NO:41.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiotensin-related disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnerrary;  
KW cell culture; chemotaxis; food additive;  
KW binding partner identification; ss.

OS Homo sapiens.



```
Query Match          100.0%; Score 72; DB 22; Length 615;
Best Local Similarity 100.0%; Pred. NO. 2.3e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atggcctctctggcctccaactgtgggtacatcctagcctctgggctttgggc 60
    |||
Db 231 atggcctctctggcctccaactgtgggtacatcctagcctctgggctttgggc 290
    |||

Oy 61 acactggtggcc 72
    |||
Db 291 acactggtggcc 302
    |||

RESULT 5
AAx97865
ID AAX97865 standard; cDNA; 1400 BP.
XX
XX AAX97865;
AC
XX 23-SEP-1999 (first entry)
XX
DE Human secreted protein encoding cDNA #53.
XX
XX Secreted protein: human; cytosolic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
XX Homo sapiens.
OS
XX
XX MO9925825-A2.
XX
XX 27-MAY-1999.
XX
XX 13-NOV-1998; 98WO-1B01862.
XX
XX 04-SEP-1998; 98US-0099273.
XX 13-NOV-1997; 97US-0066677.
XX 17-DEC-1997; 97US-0069957.
XX 09-FEB-1998; 98US-0074121.
XX 13-APR-1998; 98US-0081563.
XX 10-AUG-1998; 98US-0096116.
XX
XX (GEST ) GENSET.
XX
XX Bougueleret L, Duclet A, Dumas Milne Edwards J;
XX
XX WPI; 1999-347472/29.
XX
XX P-PSDB; AAY36181.
XX
XX Extended cDNAs encoding secreted proteins
XX
XX Claim 1; Page 254-255; 307pp; English.
XX
XX AAX97813-X97906 represent extended cDNA's which encode novel human
CC secreted proteins (see AAY36129-Y36222) and which have cytosolic,
CC thrombotic and osteopathic activity. The extended cDNAs can be used
CC to express secreted proteins or parts of them or to obtain antibodies
CC capable of binding to the secreted proteins. They may also be used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC Uses also include design of expression vectors and secretion vectors.
XX
XX Sequence 1400 BP; 326 A; 392 C; 346 G; 336 T; 0 other;
SQ
```

```
Db 96 acactggtggcc 107

RESULT 6
AAS16180
ID AAS16180 standard; cDNA; 1441 BP.
XX
XX AAS16180;
AC
XX 14-FEB-2002 (first entry)
XX
DE Human CASB81 polynucleotide.
XX
XX Human; CASB81; colorectal cancer; colon-associated disease; tumour; ss;
KW immunogen; cytosolic; antitumour; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 77..769
FT CDS /*tag= a
FT /*product= "Human CASB81 protein"
FT /*transl_except= (pos:635..637, aa:Xaa)
FT /*note= "Xaa= unknown"
XX
XX MO200180879-A2.
XX
XX 01-NOV-2001.
XX
XX 19-APR-2001; 2001WO-EP04456.
XX
XX 20-APR-2000; 2000GB-0009907.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Vinals De Bassols YC, Gaulis SRJ, Coche T, Ornloft T;
XX
XX WPI; 2002-041368/05.
XX
XX P-PSDB; AAU10521.
XX
XX Use of a claudin 2 polypeptide and polynucleotide as vaccine for
PT preventing and treating colorectal cancer, and as diagnostic reagents
PT in context of the cancer
XX
XX Claim 8; Page 57; 68pp; English.
XX
XX The invention relates to the use of CASB81 polypeptides and
CC polynucleotides for the manufacture of a medicament for treating a
CC patient suffering from or susceptible to colorectal cancer or other
CC colon-associated tumours or diseases. The sequences act as agents for
CC specific prophylactic or therapeutic immunisation against tumours because
CC they are specifically expressed or highly over-expressed in colorectal
CC tumours compared to normal cells, and can thus be targeted by
CC antigen-specific immune mechanisms leading to the destruction of tumour
CC cells. They can be used in development of an antibody immunospecific for
CC CASB81 protein, which can be used for diagnosis of tumour cells. Peptides
CC incorporating epitopes of CASB81 provide suitable immunogens. Presence of
CC or susceptibility to colorectal cancer can be diagnosed by analysing the
CC presence or amount of CASB81 DNA expression or CASB81 protein activity.
CC This sequence represents a cDNA encoding a human CASB81 polypeptide.
XX
XX Sequence 1441 BP; 337 A; 399 C; 357 G; 345 T; 3 other;
SQ
```

```
Query Match          100.0%; Score 72; DB 24; Length 1441;
Best Local Similarity 100.0%; Pred. NO. 2.8e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atggcctctctggcctccaactgtgggtacatcctagcctctgggctttgggc 60
    |||
Db 77 atggcctctctggcctccaactgtgggtacatcctagcctctgggctttgggc 136
    |||

Oy 61 acactggtggcc 72
```

```
Db 137 acactggtgcc 148
|||||
RESULT 7
AAA37060
ID AAA37060 standard; cDNA; 1475 BP.
XX
AC AAA37060;
XX
XX 08-AUG-2000 (first entry)
XX
DE Human PRO1356 (UNQ705) cDNA sequence SEQ ID NO:133.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX
XX Homo sapiens.
XX
XX WO200012708-A2.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
XX 01-SEP-1998; 98US-0098716.
XX 01-SEP-1998; 98US-0098749.
XX 01-SEP-1998; 98US-0098750.
XX 02-SEP-1998; 98US-0098803.
XX 02-SEP-1998; 98US-0098821.
XX 02-SEP-1998; 98US-0098843.
XX 09-SEP-1998; 98US-0099536.
XX 09-SEP-1998; 98US-0099596.
XX 09-SEP-1998; 98US-0099598.
XX 09-SEP-1998; 98US-0099602.
XX 09-SEP-1998; 98US-0099642.
XX 10-SEP-1998; 98US-0099741.
XX 10-SEP-1998; 98US-0099754.
XX 10-SEP-1998; 98US-0099763.
XX 10-SEP-1998; 98US-0099792.
XX 10-SEP-1998; 98US-0099808.
XX 10-SEP-1998; 98US-0099812.
XX 10-SEP-1998; 98US-0099815.
XX 15-SEP-1998; 98US-0099816.
XX 15-SEP-1998; 98US-0100385.
XX 15-SEP-1998; 98US-0100388.
XX 15-SEP-1998; 98US-0100390.
XX 16-SEP-1998; 98US-0100384.
XX 16-SEP-1998; 98US-0100627.
XX 16-SEP-1998; 98US-0100661.
XX 16-SEP-1998; 98US-0100662.
XX 16-SEP-1998; 98US-0100664.
XX 17-SEP-1998; 98US-0100683.
XX 17-SEP-1998; 98US-0100684.
XX 17-SEP-1998; 98US-0100710.
XX 17-SEP-1998; 98US-0100711.
XX 17-SEP-1998; 98US-0100919.
XX 17-SEP-1998; 98US-0100930.
XX 18-SEP-1998; 98US-0100848.
XX 18-SEP-1998; 98US-0100849.
XX 18-SEP-1998; 98US-0101014.
XX 18-SEP-1998; 98US-0101068.
XX 18-SEP-1998; 98US-0101071.
XX 22-SEP-1998; 98US-0101279.
XX 23-SEP-1998; 98US-0101471.
XX 23-SEP-1998; 98US-0101472.
XX 23-SEP-1998; 98US-0101474.
XX 23-SEP-1998; 98US-0101475.
XX 23-SEP-1998; 98US-0101476.
XX 23-SEP-1998; 98US-0101477.
XX 23-SEP-1998; 98US-0101479.
XX 24-SEP-1998; 98US-0101738.
XX 24-SEP-1998; 98US-0101741.
XX 24-SEP-1998; 98US-0101743.
XX 24-SEP-1998; 98US-0101915.
XX 24-SEP-1998; 98US-0101916.
XX 29-SEP-1998; 98US-0102207.
XX 29-SEP-1998; 98US-0102240.
XX 29-SEP-1998; 98US-0102307.
XX 29-SEP-1998; 98US-0102330.
XX 29-SEP-1998; 98US-0102331.
XX 30-SEP-1998; 98US-0102484.
XX 30-SEP-1998; 98US-0102487.
XX 30-SEP-1998; 98US-0102570.
XX 30-SEP-1998; 98US-0102571.
XX 01-OCT-1998; 98US-0102684.
XX 01-OCT-1998; 98US-0102687.
XX 02-OCT-1998; 98US-0102965.
XX 06-OCT-1998; 98US-0103258.
XX 06-OCT-1998; 98US-0103449.
XX 07-OCT-1998; 98US-0103314.
XX 07-OCT-1998; 98US-0103315.
XX 07-OCT-1998; 98US-0103328.
XX 07-OCT-1998; 98US-0103395.
XX 07-OCT-1998; 98US-0103396.
XX 07-OCT-1998; 98US-0103401.
XX 08-OCT-1998; 98US-0103633.
XX 08-OCT-1998; 98US-0103678.
XX 08-OCT-1998; 98US-0103679.
XX 08-OCT-1998; 98US-0103711.
XX 14-OCT-1998; 98US-0104257.
XX 20-OCT-1998; 98US-0104987.
XX 20-OCT-1998; 98US-0105000.
XX 20-OCT-1998; 98US-0105002.
XX 21-OCT-1998; 98US-0105104.
XX 22-OCT-1998; 98US-0105169.
XX 22-OCT-1998; 98US-0105266.
XX 26-OCT-1998; 98US-0105693.
XX 26-OCT-1998; 98US-0105694.
XX 27-OCT-1998; 98US-0105807.
XX 27-OCT-1998; 98US-0105881.
XX 27-OCT-1998; 98US-0105882.
XX 27-OCT-1998; 98US-0106062.
XX 28-OCT-1998; 98US-0106023.
XX 28-OCT-1998; 98US-0106029.
XX 28-OCT-1998; 98US-0106030.
XX 28-OCT-1998; 98US-0106032.
XX 28-OCT-1998; 98US-0106033.
XX 28-OCT-1998; 98US-0106178.
XX 29-OCT-1998; 98US-0106248.
XX 29-OCT-1998; 98US-0106384.
XX 30-OCT-1998; 98US-0108500.
XX 30-OCT-1998; 98US-0108506.
XX 03-NOV-1998; 98US-0106856.
XX 03-NOV-1998; 98US-0106902.
XX 03-NOV-1998; 98US-0106905.
XX 03-NOV-1998; 98US-0106919.
XX 03-NOV-1998; 98US-0106932.
XX 03-NOV-1998; 98US-0106934.
XX 10-NOV-1998; 98US-0107783.
XX 17-NOV-1998; 98US-0108775.
XX 17-NOV-1998; 98US-0108779.
XX 17-NOV-1998; 98US-0108787.
XX 17-NOV-1998; 98US-0108788.
XX 17-NOV-1998; 98US-0108801.
XX 17-NOV-1998; 98US-0108802.
XX 17-NOV-1998; 98US-0108806.
XX 17-NOV-1998; 98US-0108807.
XX 17-NOV-1998; 98US-0108867.
XX 17-NOV-1998; 98US-0108925.
XX 18-NOV-1998; 98US-0108848.
XX 18-NOV-1998; 98US-0108849.
XX 18-NOV-1998; 98US-0108850.
XX 18-NOV-1998; 98US-0108851.
XX 18-NOV-1998; 98US-0108852.
```



KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX Homo sapiens.

OS WO200140466-A2.

XX 07-JUN-2001.

XX 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

XX 01-DEC-1999; 99WO-US28634.

XX 02-DEC-1999; 99WO-US28551.

XX 02-DEC-1999; 99WO-US28564.

XX 02-DEC-1999; 99WO-US28565.

XX 09-DEC-1999; 99US-0170262.

XX 16-DEC-1999; 99WO-US30095.

XX 20-DEC-1999; 99WO-US30911.

XX 20-DEC-1999; 99WO-US30999.

XX 30-DEC-1999; 99WO-US31243.

XX 06-JAN-2000; 2000WO-US00277.

XX 06-JAN-2000; 2000WO-US00376.

XX 11-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 24-FEB-2000; 2000WO-US04914.

XX 24-FEB-2000; 2000WO-US05004.

XX 01-MAR-2000; 2000WO-US05601.

XX 20-MAR-2000; 2000WO-US07377.

XX 21-MAR-2000; 2000WO-US07532.

XX 30-MAR-2000; 2000WO-US08439.

XX 17-MAY-2000; 2000WO-US13705.

XX 22-MAY-2000; 2000WO-US14042.

XX 30-MAY-2000; 2000WO-US14941.

XX 02-JUN-2000; 2000WO-US15264.

XX 10-NOV-2000; 2000WO-US30873.

XX (GETH ) GENENTECH INC.

XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

XX Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

XX Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX WPI; 2001-408281/43.

XX P-PSDB; AAU12417.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect

XX other PRO polypeptides, link bioactive molecules to cells expressing

XX PRO polypeptides, and detect the presence of mammalian tumours e.g.

XX lung, breast, prostate, cervical

XX Claim 3; Fig 491; 813pp; English.

XX AAS21244-AAS21518 encode for novel human secretory and transmembrane

XX PRO polypeptides. The PRO polypeptides are useful to detect other

XX PRO polypeptides, to link bioactive molecules to cells expressing

XX PRO polypeptides, to modulate biological activities of cells expressing

XX PRO polypeptides, and to detect the presence of mammalian lung, colon,

XX breast, prostate, rectal, cervical or liver tumours by comparing PRO

XX polypeptide expression in a cell sample to that in a control sample.

XX Some of the 275 sequences are also useful to stimulate the release of

XX tumour necrosis factor-alpha (TNF-alpha) from human blood, the

XX proliferation or differentiation of chondrocytes, the proliferation or

XX gene expression in pericyte cells, the release of proteoglycans from

XX cartilage, the proliferation of inner ear utricular supporting cells or

XX monocytes (PBMCs), the release of a cytokine from peripheral blood

XX the PRO polypeptides may modulate glucose or free fatty acid uptake by

XX skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide

XX to factor VIIA. The PRO polypeptides can be used in assays to identify

XX molecules involved in binding interactions. The polynucleotides encoding

XX PRO polypeptides can be used to generate probes, antisense RNA/DNA,

CC transgenic or knock out animals and can be used in gene therapy.

XX Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

Query Match 100.0%; Score 72; DB 22; Length 1475;

Best Local Similarity 100.0%; Pred. No. 2.9e-15;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctccaaactgtggctacatctagcctctctgtggccttttgggc 60

Db 122 atggcctctcttgccctccaaactgtggctacatctagcctctctgtggccttttgggc 181

QY 61 acactgggttgc 72

Db 182 acactgggttgc 193

RESULT 10

AAF92097

ID AAF92097 standard; cDNA; 1475 BP.

XX AAF92097;

DT 15-MAY-2001 (first entry)

XX Human PRO1356 cDNA.

XX Human; PRO protein; mapping; ss.

XX Homo sapiens.

XX WO200116318-A2.

PD 08-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 09-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GETH ) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2001-183260/18.

XX P-PSDB; AAB87565.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

XX molecular biology, including use as hybridization probes, and in

XX chromosome and gene mapping.

XX Claim 2; Fig 79; 278pp; English.

XX The present sequence is the coding sequence for a human PRO polypeptide

XX (secreted and transmembrane). The PRO protein, and PRO agonists, PRO

XX antagonists or anti-PRO antibodies are useful for preparation of a

XX medicament useful in the treatment of a condition which is responsive to

XX the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO

XX protein may also be employed as molecular weight markers for protein

XX electrophoresis. The PRO coding sequence has applications in molecular

CC biology, including use as hybridisation probes, and in chromosome and  
CC gene mapping.

XX Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

XX Query Match 100.0%; Score 72; DB 22; Length 1475;

XX Best Local Similarity 100.0%; Pred. No. 2.9e-15;

XX Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctctctgtggcctcccaactgtggctacatccttagcctcttgaggc 60

Db 122 atggcctctctgtggcctcccaactgtggctacatccttagcctcttgaggc 181

OY 61 acactgtgtgcc 72

Db 182 acactgtgtgcc 193

RESULT 11

AAF54296 standard; DNA; 1475 BP.

XX AAF54296;

DT 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #39.

XX Secreted; transmembrane; gene therapy; ss.

OS Unidentified.

PN WO200078961-A1.

PD 28-DEC-2000.

PF 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GENTH ) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX Gao W, Goddard A, Godowski PJ, Grimaldi CD, Gurrey AL, Hillan KJ;

XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

XX Secrected and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

XX therapy -

XX Claim 2; Fig 77; 787pp; English.

XX Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

XX Query Match 100.0%; Score 72; DB 22; Length 1475;

XX Best Local Similarity 100.0%; Pred. No. 2.9e-15;

XX Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctctctgtggcctcccaactgtggctacatccttagcctcttgaggc 60

Db 122 atggcctctctgtggcctcccaactgtggctacatccttagcctcttgaggc 181

OY 61 acactgtgtgcc 72

Db 182 acactgtgtgcc 193

RESULT 12

AAD08497 standard; cDNA; 1481 BP.

XX AAD08497;

DT 09-AUG-2001 (first entry)

DE Human secreted protein-encoding gene 10 cDNA clone HRP183, SEQ ID NO:20.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

XX foetal abnormality; developmental abnormality; haematopoietic disorder;

XX immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;

XX inflammation; allergy; neurological disorder; Alzheimer's disease;

XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;

XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

XX cardiovascular disorder; angioecnic disorder; kidney disorder;

XX gastrointestinal disorder; pregnancy-related disorder; gene therapy;

XX endocrine disorder; infection; wound healing; vulnerraty;

XX cell culture; chemotaxis; food additive;

XX binding partner identification; ss.

OS Homo sapiens.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US31162.

XX 19-NOV-1999; 99US-0166415.

XX 30-JUN-2000; 2000US-0215136.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;

XX WPI; 2001-343793/36.

XX P-PSDB; AAE04207.

XX Isolated nucleic acid molecule encoding a human secreted protein is

XX used in preventing, treating or ameliorating a medical condition -

XX Claim 1; Page 410; 509pp; English.

XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted

XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.

CC AAE04240-AAE04297 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing, gene  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 18 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin ageing due to  
CC sunburn, and maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.  
XX  
SQ Sequence 1481 BP; 353 A; 405 C; 375 G; 348 T; 0 other;

Query Match 100.0%; Score 72; DB 22; Length 1481;  
Best Local Similarity 100.0%; Pred. No. 2.9e-15;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctctcttggtcccaactgtgggtacatcctagcctcttctgggcttttgggc 60  
|||||  
DB 118 atggcctctcttggtcccaactgtgggtacatcctagcctcttctgggcttttgggc 177  
|||||  
QY 61 acactggttgcc 72  
|||||  
DB 178 acactggttgcc 189

RESULT 13  
AAx97818  
ID AAX97818 standard; cDNA; 1524 BP.  
XX  
AC AAX97818;  
XX  
DT 23-SEP-1999 (first entry)  
XX  
DE Human secreted protein encoding cDNA #6.  
XX  
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;  
KW diagnostic; gene therapy; chromosome mapping; secretion vector; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO925825-A2.  
XX  
PD 27-MAY-1999.  
XX  
PF 13-NOV-1998; 98WO-IB01862.  
XX  
PR 04-SEP-1998; 98US-0099273.  
PR 13-NOV-1997; 97US-0066677.  
PR 17-DEC-1997; 97US-0069957.  
PR 09-FEB-1998; 98US-0074121.  
PR 13-APR-1998; 98US-0081563.  
PR 10-AUG-1998; 98US-0096116.  
XX  
PA (GEST ) GENSET.

XX Bougueleret L, Duclert A, Dumas Milne Edwards J;  
XX  
DR WPI; 1999-347472/29.  
DR P-PSDB; AAY36134.  
XX  
PT Extended cDNAs encoding secreted proteins  
XX  
PS Example 28; Page 155-156; 307pp; English.  
XX  
CC AAX97813-X97906 represent extended cDNA's which encode novel human  
CC secreted proteins (see AAY36129-V36222) and which have cytostatic, to  
CC thrombotic and osteopathic activity. The extended cDNAs can be used to  
CC express secreted proteins or parts of them or to obtain antibodies  
CC capable of binding to the secreted proteins. They may also be used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC Uses also include design of expression vectors and secretion vectors.  
XX  
SQ Sequence 1524 BP; 365 A; 404 C; 382 G; 353 T; 20 other;

Query Match 100.0%; Score 72; DB 20; Length 1524;  
Best Local Similarity 100.0%; Pred. No. 2.9e-15;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctctcttggtcccaactgtgggtacatcctagcctcttctgggcttttgggc 60  
|||||  
DB 160 atggcctctcttggtcccaactgtgggtacatcctagcctcttctgggcttttgggc 219  
|||||  
QY 61 acactggttgcc 72  
|||||  
DB 220 acactggttgcc 231

RESULT 14  
AAI58013  
ID AAI58013 standard; cDNA; 1540 BP.  
XX  
AC AAI58013;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 216.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

```
XX DR WPI: 2001-442253/47.
XX P-PSDB; AAM38857.
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
PS Claim 1; SEQ ID NO 216; 10078bp; English.
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX SQ Sequence 1540 BP; 365 A; 419 C; 389 G; 367 T; 0 other;
Query Match 100.0%; Score 72; DB 22; Length 1540;
Best Local Similarity 100.0%; Pred. No. 2,9e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggcctctcttggtcccaactgtgggtacacccatagcctcttggtggtttgggc 60
DB 183 atggcctctcttggtcccaactgtgggtacacccatagcctcttggtggtttgggc 242
QY 61 acactggttggc 72
DB 243 acactggttggc 254
RESULT 15
AA159799/c
ID AA159799 standard; cDNA; 1761 BP.
XX
XX AC AA159799;
XX
XX DT 22-OCT-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 3788.
XX
XX KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000MO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
```

```
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou F, Goodrich R, Drmanac RT;
XX
XX DR WPI: 2001-442253/47.
XX P-PSDB; AAM40643.
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
PS Claim 1; SEQ ID NO 3788; 10078bp; English.
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX SQ Sequence 1761 BP; 407 A; 446 C; 478 G; 430 T; 0 other;
Query Match 100.0%; Score 72; DB 22; Length 1761;
Best Local Similarity 100.0%; Pred. No. 3e-15;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggcctctcttggtcccaactgtgggtacacccatagcctcttggtggtttgggc 60
DB 1641 atggcctctcttggtcccaactgtgggtacacccatagcctcttggtggtttgggc 1582
QY 61 acactggttggc 72
DB 1581 ACACCTGCTTGGC 1570
```

Search completed: June 2, 2002, 20:09:03  
Job time: 14748 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:19:26 ; Search time 118.89 Seconds  
(without alignments)  
214.879 Million cell updates/sec

Title: US-09-663-600a-186  
Perfect score: 1174  
Sequence: 1 MASIGLQLVYLGLLGLG.....PCQPPKVKSEFNSTLTGIV 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1174	100.0	230	AAV36181	Human secreted pro
2	1174	100.0	230	AAV99378	Human PRO1356 (UNQ
3	1174	100.0	230	AAV84609	A human membrane a
4	1174	100.0	230	AAU09178	Human PRO1356 poly
5	1174	100.0	230	AAU12417	Human PRO1356 poly
6	1174	100.0	230	AAE04207	Human gene 10 enco
7	1174	100.0	230	AAE87565	Human PRO1356. Ho
8	1174	100.0	230	AAE88342	Human membrane or
9	1174	100.0	230	AAE66127	Protein of the inv
10	1169	99.6	230	AAV38857	Human polypeptide
11	1169	99.6	260	AAV40643	Human polypeptide

12	1156	98.5	230	20	AAV36134	Human secreted pro
13	1096	93.4	230	21	AAV51676	Murine clodin 2 pr
14	695	59.2	140	22	AAE04228	Human gene 10 enco
15	635	54.1	155	21	AAV54052	Human pancreatic c
16	571	48.6	239	21	AAV99431	Human PRO1571 (UNQ
17	571	48.6	239	22	AAU29200	Human PRO polypept
18	571	48.6	239	22	AAE66180	Protein of the inv
19	570	48.6	113	22	AAE04281	Human gene 10 enco
20	559	47.6	114	20	AAV12226	Human 5' EST secre
21	529	45.1	233	22	AAV78559	Human protein SEQ
22	515	43.9	233	22	AAV79543	Human protein SEQ
23	455	38.8	210	20	AAW86307	Kidney injury asso
24	450	38.3	220	20	AAV13939	Human transmembran
25	450	38.3	220	22	ABE50279	Claudin 3 ovarian
26	448	38.2	88	22	AAV75581	Gene 48 human secr
27	446	38.0	208	21	AAV52100	Murine clodin 6 pr
28	443	37.7	219	21	AAV51681	Human ORFX ORF2857
29	442.5	37.7	209	21	AAV43133	Claudin 4 ovarian
30	442.5	37.7	209	22	ABE50278	Human colon cancer
31	442.5	37.7	210	22	AAV75467	Amino acid sequenc
32	441	37.6	217	22	AAV64401	Murine clodin 3 pr
33	441	37.6	219	21	AAV51679	Human polypeptide
34	438	37.3	220	21	AAV99434	Human membrane or
35	438	37.3	220	22	AAE66183	Human polypeptide
36	438	37.3	220	22	AAE50968	Novel human diagno
37	437	37.2	220	22	AAV39937	Murine clodin 7 pr
38	437	37.2	220	22	AAE88319	Human gene 10 enco
39	437	37.2	235	22	AAV41723	Human secreted pro
40	434.5	37.0	254	22	ABG27980	Human secreted pro
41	431.5	36.8	211	21	AAV51682	Human apoptosis as
42	430	36.6	86	22	AAE04229	
43	427.5	36.4	205	21	AAV52101	
44	427.5	36.4	210	19	AAV63697	
45	427.5	36.4	211	21	AAV57609	

ALIGNMENTS

RESULT 1  
AAV36181  
ID AAV36181 standard; protein; 230 AA.  
XX  
AC AAV36181;  
XX  
DT 23-SEP-1999 (first entry)  
XX  
DE Human secreted protein #53.  
XX  
KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;  
KW diagnostic; gene therapy; chromosome mapping; secretion vector.  
XX  
OS Homo sapiens.  
XX  
PN WO9925825-A2.  
XX  
PD 27-MAY-1999.  
XX  
PF 13-NOV-1998; 98WO-TB01862.  
XX  
PR 04-SEP-1998; 98US-0099273.  
PR 13-NOV-1997; 97US-0066677.  
PR 17-DEC-1997; 97US-0069957.  
PR 09-FEB-1998; 98US-0074121.  
PR 13-APR-1998; 98US-0081563.  
PR 10-AUG-1998; 98US-0096116.  
PA (GEST ) GENSET.  
XX  
PI Bouqueleret L, Duclert A, Dumas Milne Edwards J;  
XX WPI; 1999-347472/29.  
XX DR N-PSDB; AAV97865.

XX Extended cDNAs encoding secreted proteins  
XX Claim 7; Page 289; 307pp: English.  
XX AAY36129-v36222 represent novel human secreted proteins encoded by the  
CC extended cDNA sequences represented in AAY97813-X97906. The proteins  
CC of the invention have cytosolic, thrombotic and osteopathic activity.  
CC The extended cDNAs can be used to express secreted proteins or parts of  
CC them or to obtain antibodies capable of binding to the secreted  
CC proteins. They may also be used in diagnostic, forensic, gene therapy  
CC and chromosome mapping procedures. Uses also include design of  
CC expression vectors and secretion vectors.  
XX  
SQ Sequence 230 AA:  
  
Query Match 100.0%; Score 1174; DB 20; Length 230;  
Best Local Similarity 100.0%; Pred. No. 4, 1e-117;  
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MASIGQLVGYILGLGLTIVAMLPMSKTSYGVASIVTAVGSKGLMECATSTG 60  
Db 1 masigqlvgyilglgltlvamlpswktsygvastvavgfkskglmecatstg 60  
QY 61 ITQCDDYITLIGLPADIQAAQAMMTSSAITSACITSVGMCTVPCQESRAKRVAVA 120  
Db 61 itqcdylstlilgipadiqaqammvtsaistacitstvgmctvcqesrakdrvava 120  
QY 121 GGVFFILGILGIFPVAWNHGILRDYSPILVDPMSKFEIGEMALYIGITISLFLGII 180  
Db 121 ggvffilgllglfpwamnhgilrdysplvpmskfeligemalyigistlsflgii 180  
QY 181 LCFSCSSQRNRNRYDAYQAQAPLATRSSPRGQPKYKSEFNSTLNGYV 230  
Db 181 lcfscssqnrnsnydayaqaplatrssprgqpkvksefnstlgyv 230  
  
RESULT 2  
AAY9378  
ID AAY9378 standard; Protein; 230 AA.  
XX  
AC AAY9378;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Human PRO1356 (UNQ705) amino acid sequence SEQ ID NO:134.  
XX  
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
KW Transmembrane; secretion; Immunoadhesion; Pharmaceutical; screening.  
XX  
OS Homo sapiens.  
XX  
PN WO200012708-A2.  
XX  
PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US20111.  
XX  
PR 01-SEP-1998; 98US-0098716.  
PR 01-SEP-1998; 98US-0098749.  
PR 01-SEP-1998; 98US-0098750.  
PR 02-SEP-1998; 98US-0098803.  
PR 02-SEP-1998; 98US-0098821.  
PR 02-SEP-1998; 98US-0098843.  
PR 09-SEP-1998; 98US-0099336.  
PR 09-SEP-1998; 98US-0099596.  
PR 09-SEP-1998; 98US-0099598.  
PR 09-SEP-1998; 98US-0099602.  
PR 09-SEP-1998; 98US-0099642.  
PR 10-SEP-1998; 98US-0099741.  
PR 10-SEP-1998; 98US-0099754.  
PR 10-SEP-1998; 98US-0099763.

PR 10-SEP-1998; 98US-0099792.  
PR 10-SEP-1998; 98US-0099808.  
PR 10-SEP-1998; 98US-0099812.  
PR 10-SEP-1998; 98US-0099815.  
PR 10-SEP-1998; 98US-0099816.  
PR 15-SEP-1998; 98US-0100365.  
PR 15-SEP-1998; 98US-0100388.  
PR 15-SEP-1998; 98US-0100390.  
PR 16-SEP-1998; 98US-0100584.  
PR 16-SEP-1998; 98US-0100627.  
PR 16-SEP-1998; 98US-0100661.  
PR 16-SEP-1998; 98US-0100662.  
PR 16-SEP-1998; 98US-0100664.  
PR 17-SEP-1998; 98US-0100683.  
PR 17-SEP-1998; 98US-0100684.  
PR 17-SEP-1998; 98US-0100710.  
PR 17-SEP-1998; 98US-0100711.  
PR 17-SEP-1998; 98US-0100919.  
PR 17-SEP-1998; 98US-0100930.  
PR 18-SEP-1998; 98US-0100848.  
PR 18-SEP-1998; 98US-0100849.  
PR 18-SEP-1998; 98US-0101014.  
PR 18-SEP-1998; 98US-0101016.  
PR 18-SEP-1998; 98US-0101068.  
PR 22-SEP-1998; 98US-0101071.  
PR 22-SEP-1998; 98US-0101279.  
PR 23-SEP-1998; 98US-0101471.  
PR 23-SEP-1998; 98US-0101472.  
PR 23-SEP-1998; 98US-0101474.  
PR 23-SEP-1998; 98US-0101475.  
PR 23-SEP-1998; 98US-0101476.  
PR 23-SEP-1998; 98US-0101477.  
PR 23-SEP-1998; 98US-0101479.  
PR 24-SEP-1998; 98US-0101738.  
PR 24-SEP-1998; 98US-0101741.  
PR 24-SEP-1998; 98US-0101743.  
PR 24-SEP-1998; 98US-0101915.  
PR 24-SEP-1998; 98US-0101916.  
PR 29-SEP-1998; 98US-0102207.  
PR 29-SEP-1998; 98US-0102240.  
PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102330.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.  
PR 30-SEP-1998; 98US-0102570.  
PR 30-SEP-1998; 98US-0102571.  
PR 01-OCT-1998; 98US-0102684.  
PR 01-OCT-1998; 98US-0102687.  
PR 02-OCT-1998; 98US-0102965.  
PR 06-OCT-1998; 98US-0103258.  
PR 06-OCT-1998; 98US-0103449.  
PR 07-OCT-1998; 98US-0103314.  
PR 07-OCT-1998; 98US-0103315.  
PR 07-OCT-1998; 98US-0103328.  
PR 07-OCT-1998; 98US-0103395.  
PR 07-OCT-1998; 98US-0103396.  
PR 07-OCT-1998; 98US-0103401.  
PR 08-OCT-1998; 98US-0103633.  
PR 08-OCT-1998; 98US-0103678.  
PR 08-OCT-1998; 98US-0103679.  
PR 08-OCT-1998; 98US-0103711.  
PR 14-OCT-1998; 98US-0104257.  
PR 20-OCT-1998; 98US-0104987.  
PR 20-OCT-1998; 98US-0105000.  
PR 20-OCT-1998; 98US-0105002.  
PR 21-OCT-1998; 98US-0105104.  
PR 22-OCT-1998; 98US-0105169.  
PR 22-OCT-1998; 98US-0105266.  
PR 26-OCT-1998; 98US-0105693.  
PR 26-OCT-1998; 98US-0105694.  
PR 27-OCT-1998; 98US-0105807.  
PR 27-OCT-1998; 98US-0105881.  
PR 27-OCT-1998; 98US-0105882.

```
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0108905.
PR 03-NOV-1998; 98US-0108919.
PR 03-NOV-1998; 98US-0108932.
PR 03-NOV-1998; 98US-0108934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX
FA (GETH ) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
XX WPI: 2000-237871/20.
DR N-PSDB; AAA37060.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 78; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding them have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
XX Sequence 230 AA;
SQ

Query Match 100.08; Score 1174; DB 21; Length 230;
Best Local Similarity 100.08; Pred. No. 4.1e-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLGLQLVGYILGLLGLTGLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHTSG 60
DB 1 maslglqlvgyilgllglgtivamlpsswktssyvgasivtavgfsgklwmeccathstg 60
QY 61 ITQCDIYSTLLGLPADIQAAQAMMTYSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120
DB 61 itqcdiystllglpadiaqaqamvmvtssaislaciisvvgmrctvfcqesrakdrava 120

QY 121 GGVFFILGGLGIFIPVANNLHGILRDFYSPLPDMSMKFEIGALYLGIISSLSIAGII 180
DB 121 ggvffilgglgifipvawnlhgilrdfysplpdsmskfeigalylgliisslfsliagi 180
QY 181 LCFSCSSORNSRNYDAYOAOPLATRSSPRGPPKVKSEFNSYSLTGYV 230
DB 181 lcfscssqrnsrnydayaqagqatrrsprp9qppkvksefnysltgyv 230

RESULT 3
AAY84609
ID AAY84609 standard; Protein; 230 AA.
XX
AC AAY84609;
XX
XX 25-JUL-2000 (first entry)
XX
DE A human membrane associated organizational protein (HJUNCT).
XX
KW Human; membrane associated organizational protein; HJUNCT;
KW cell proliferative disorder; cancer; autoimmune disorder;
KW inflammatory disorder; neurological disorder; developmental disorder;
KW vesicle trafficking; reproductive disorder; gastrointestinal disorder;
KW renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;
KW Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;
KW irritable bowel syndrome; allergy.
XX
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 29
FT FT /note= "potential phosphorylation site"
FT Modified-site 62
FT FT /note= "potential phosphorylation site"
FT Domain 117..138
FT FT /note= "transmembrane domain"
FT Modified-site 155
FT FT /note= "potential phosphorylation site"
FT Domain 164..182
FT FT /note= "transmembrane domain"
FT Modified-site 187
FT FT /note= "potential phosphorylation site"
FT Modified-site 190
FT FT /note= "potential glycosylation site"
FT Modified-site 208
FT FT /note= "potential phosphorylation site"
FT Modified-site 224
FT FT /note= "potential phosphorylation site"
XX
XX WO200018915-A2.
XX
XX 06-APR-2000.
XX
XX 23-SEP-1999; 99WO-US22082.
XX
XX 25-SEP-1998; 98US-0155215.
XX 13-OCT-1998; 98US-0155251.
XX 04-MAY-1999; 99US-0172228.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR, Lu AD, Tang YT;
XX
XX WPI: 2000-293154/25.
XX N-PSDB; AAA12585.
XX
XX Human membrane associated organizational protein and nucleic acid
PT sequences useful in the diagnosis, treatment and prevention of cell
PT proliferative associated disorders e.g. cancer, rheumatoid arthritis
PT and Alzheimer's disease -
XX
XX Claim 1; Page 76-77; 84pp; English.
PS
```



```
SQ      Sequence      230 AA;

Query Match      100.0%; Score 1174; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.le-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASIGLQLVGYILGLLGLTGLVAMLLPSWKTSYVGVASIVTAVGFSKGLWMECATHSTG 60
      |||||||
DB      1 maslglqlvgylglglgtlvalmlpswktssyvgasivtavgfskglwmcathstg 60

QY      61 ITQCDIYSTLLGLPADIOAAQAMVTSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120
      |||||||
DB      61 itqcdiystllglpadiqaqaamvtssaislaciisvvgmrctvcqesrakdrvava 120

QY      121 GGVEFFILGGLGFIYVAVNLHGILRDFYSPVLPDSMKFEIGEALYLGIISSLSFIAGI 180
      |||||||
DB      121 ggveffilgglgfiypawnlhgilrdfsplvpdsmkfeigealylgiisslsfiagi 180

QY      181 LCFSCSSQRNSRNYDAYQAOPLATRRSPRPGQPKVKSEFNSYSLTGYV 230
      |||||||
DB      181 lcfscssqrnsnydayqaoqlatrrsprpgqpkvksefnsyltgyv 230

RESULT      5
AAU12417
ID      AAU12417 standard; Protein; 230 AA.
AC
XX
XX
XX      24-OCT-2001 (first entry)
XX
DE      Human PRO1356 polypeptide sequence.
XX
KW      Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW      breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW      cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW      adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS      Homo sapiens.
XX
XX      WO20010466-A2.
XX
PD      07-JUN-2001.
XX
XX      01-DEC-2000; 2000WO-US32678.
XX
PR      01-DEC-1999; 99WO-US28301.
PR      01-DEC-1999; 99WO-US28634.
PR      02-DEC-1999; 99WO-US28551.
PR      02-DEC-1999; 99WO-US28564.
PR      02-DEC-1999; 99WO-US28565.
PR      09-DEC-1999; 99US-0170262.
PR      16-DEC-1999; 99WO-US30095.
PR      20-DEC-1999; 99WO-US30911.
PR      20-DEC-1999; 99WO-US30999.
PR      30-DEC-1999; 99WO-US31243.
PR      06-JAN-2000; 2000WO-US00277.
PR      06-JAN-2000; 2000WO-US00376.
PR      11-FEB-2000; 2000WO-US03565.
PR      18-FEB-2000; 2000WO-US04341.
PR      18-FEB-2000; 2000WO-US04342.
PR      22-FEB-2000; 2000WO-US04414.
PR      24-FEB-2000; 2000WO-US04914.
PR      24-FEB-2000; 2000WO-US05004.
PR      01-MAR-2000; 2000WO-US05601.
PR      20-MAR-2000; 2000WO-US07377.
PR      21-MAR-2000; 2000WO-US07532.
PR      30-MAR-2000; 2000WO-US08439.
PR      17-MAY-2000; 2000WO-US13705.
PR      22-MAY-2000; 2000WO-US14042.
PR      30-MAY-2000; 2000WO-US14941.
PR      02-JUN-2000; 2000WO-US15264.
```

```
PR      10-NOV-2000; 2000WO-US30873.
XX
PA      (GETH ) GENENTECH INC.
XX
PI      Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI      Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI      Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX      WPI; 2001-408281/43.
DR      N-PSDB; AAS21489.
XX
PT      Isolated, secretory and transmembrane PRO polypeptide used to detect
PT      other PRO polypeptides, link bioactive molecules to cells expressing
PT      PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT      lung, breast, prostate, cervical
XX
PS      Claim 12; Fig 492; 813pp; English.
XX
CC      AAU12172-AAU12446 represent novel human secretory and transmembrane
CC      PRO polypeptides. The PRO polypeptides are useful to detect other
CC      PRO polypeptides, to link bioactive molecules to cells expressing
CC      PRO polypeptides, to modulate biological activities of cells expressing
CC      PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC      breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC      polypeptide expression in a cell sample to that in a control sample.
CC      Some of the 275 sequences are also useful to stimulate the release of
CC      tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC      proliferation or differentiation of chondrocytes, the proliferation or
CC      gene expression in pericyte cells, the release of proteolytases from
CC      cartilage, the proliferation of inner ear utricular supporting cells or
CC      of T-lymphocytes, the release of a cytokine from peripheral blood
CC      monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC      the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC      skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC      to factor VIIA. The PRO polypeptides can be used in assays to identify
CC      molecules involved in binding interactions. The polynucleotides encoding
CC      PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC      transgenic or knock out animals and can be used in gene therapy.
XX
SQ      Sequence      230 AA;

Query Match      100.0%; Score 1174; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.le-117;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASIGLQLVGYILGLLGLTGLVAMLLPSWKTSYVGVASIVTAVGFSKGLWMECATHSTG 60
      |||||||
DB      1 maslglqlvgylglglgtlvalmlpswktssyvgasivtavgfskglwmcathstg 60

QY      61 ITQCDIYSTLLGLPADIOAAQAMVTSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120
      |||||||
DB      61 itqcdiystllglpadiqaqaamvtssaislaciisvvgmrctvcqesrakdrvava 120

QY      121 GGVEFFILGGLGFIYVAVNLHGILRDFYSPVLPDSMKFEIGEALYLGIISSLSFIAGI 180
      |||||||
DB      121 ggveffilgglgfiypawnlhgilrdfsplvpdsmkfeigealylgiisslsfiagi 180

QY      181 LCFSCSSQRNSRNYDAYQAOPLATRRSPRPGQPKVKSEFNSYSLTGYV 230
      |||||||
DB      181 lcfscssqrnsnydayqaoqlatrrsprpgqpkvksefnsyltgyv 230

RESULT      6
AAE04207
ID      AAE04207 standard; Protein; 230 AA.
XX
XX      AAE04207;
AC
XX
XX      09-AUG-2001 (first entry)
XX
DE      Human gene 10 encoded secreted protein HTP1H83, SEQ ID NO:62.
XX
```

KW	Human; secreted protein; proliferative disorder; cancer; tumour;
KW	fetal abnormality; developmental abnormality; haematopoietic disorder;
KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW	Inflammation; allergy; neurological disorder; Alzheimer's disease;
KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW	cardiovascular disorder; angiogenic disorder; kidney disorder;
KW	gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW	endocrine disorder; infection; wound healing; vulnerary;
KW	cell culture; chemotaxis; food additive;
KW	binding partner identification.
XX	Homo sapiens.
OS	
XX	
FH	Key
FT	Peptide
FT	Protein
FT	
PN	WO200136432-A2.
XX	
PD	25-MAY-2001.
XX	
PF	15-NOV-2000; 2000WO-US31162.
PR	19-NOV-1999; 99US-0166415.
PR	30-JUN-2000; 2000US-0215136.
XX	
PA	(HOMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Komatsoulis GA, Baker KP, Young PE;
DR	WPt: 2001-343793/36.
DR	N-PSDB; AAD08497.
PT	
PT	Isolated nucleic acid molecule encoding a human secreted protein is
PS	used in preventing, treating or ameliorating a medical condition -
XX	
XX	Claim 11; Page 439; 509pp: English.
CC	AAD08488-AA008529 represent cDNAs corresponding to 18 human secreted
CC	protein genes, and AAE04199-AAE04239 represent the proteins they encode,
CC	AAE04240-AAE04297 represent human secreted protein fragments or variants
CC	The secreted proteins and their genes are useful for preventing,
CC	treatment or ameliorating medical conditions, e.g., by protein or gene
CC	therapy. Pathological conditions can be diagnosed by determining the
CC	amount of the new protein in a sample or by determining the presence of
CC	mutations in the new genes. Specific uses are described for each of the
CC	18 genes, based on the tissues in which they are most highly expressed,
CC	and include developing products for the diagnosis or treatment of
CC	proliferative disorders, cancer, tumours, foetal and developmental
CC	abnormalities, haematopoietic disorders, diseases of the immune system,
CC	AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC	allergies, neurological disorders (e.g., Alzheimer's disease,
CC	Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC	skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC	cardiovascular disorders, angiocenic disorders, kidney disorders,
CC	gastrointestinal disorders, pregnancy-related disorders, endocrine
CC	disorders, and infections. The proteins can also be used to aid wound
CC	healing and epithelial cell proliferation, to prevent skin ageing due to
CC	sunburn, to maintain organs before transplantation, for supporting cell
CC	culture of primary tissues, to regenerate tissues, to identify their
CC	cognate ligands or binding partners, and in chemotaxis, and can be used
CC	as a food additive or preservative to modify storage properties.
CC	Antibodies specific for a protein of the invention can be used in
CC	alleviating symptoms associated with the disorders mentioned above, and
CC	in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC	immunosorbent assay (ELISA). The present sequence represents a human
CC	secreted protein of the invention.
XX	
Sequence	230 AA;

	Query Match	100.0%	Score 1174	DB 22	Length 230	
	Best Local Similarity	100.0%	Pred. No. 4	1e-117		
	Matches 230	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
Qy	1 MASIQLQVLYGILGLGLTGLVAMLLPSWKTSSYVGASIVTAVGFSKGLMECATHTSG	60				
Db	1 masilqlqvlylglllgllgtlvamllpswktsyvgasivtavgfskglmwecathstg	60				
Qy	61 ITQCDIVSTLGLPADIDQAQAMMTSSAISLACIIISVGNKRCIVFCQESAKRNVAYA	120				
Db	61 itqcdivstllglpadidqaagmmvltssaislaciisvgnrcvfcqesakrdnvaya	120				
Qy	121 GGVEFIILGLGFIIVANMLHGLRDFYSPLYPDSMKFEIGALYLGIIISLFLIAGII	180				
Db	121 ggvefiiilgllgfiivawnllhgilrdfsplypdsmkfciqgalylgiiisflsilaagii	180				
Qy	181 LCFSCSSQRNMSNYDAYOAPLAFTRSSPRGQPRKVSSEFNSSYSLTGTV	230				
Db	181 lcfsscsgqrnmsnydayaqaplatrssprpqpkkvksfnssyltgyv	230				
RESULT 7						
AA87565	AA87565 standard; Protein, 230 AA.					
AA87565	AA87565;					
DT	15-MAY-2001 (first entry)					
DE	Human PRO1356.					
KW	Human; PRO protein; mapping.					
OS	Homo sapiens.					
PN	WO200116318-A2.					
PD	08-MAR-2001.					
PE	24-AUG-2000; 2000WO-US23328.					
PR	01-SEP-1999; 99WO-US20111.					
PR	15-SEP-1999; 99WO-US21090.					
PR	07-DEC-1999; 99US-0169495.					
PR	09-DEC-1999; 99US-0170262.					
PR	11-JAN-2000; 2000US-0175481.					
PR	18-FEB-2000; 2000WO-US04341.					
PR	18-FEB-2000; 2000WO-US04342.					
PR	22-FEB-2000; 2000WO-US04414.					
PR	01-MAR-2000; 2000WO-US05601.					
PR	03-MAR-2000; 2000US-0187202.					
PR	25-APR-2000; 2000US-0199397.					
PR	22-MAY-2000; 2000WO-US14042.					
PR	05-JUN-2000; 2000US-0209832.					
PA	(GENE ) GENENTECH INC.					
PI	Eaton DL, Filvarcoff E, Gerlitsen ME, Goddard A, Godowski PJ;					
PI	Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;					
XX	WPI: 2001-183260/18.					
DR	N-PSDB; AAF92097.					
PT	Eighty four nucleic acids encoding PRO polypeptides, useful in					
PT	molecular biology, including use as hybridization probes, and in					
XX	chromosome and gene mapping. -					
XX	Claim 12; Fig 80; 278pp; English.					
CC	The present sequence is a human PRO polypeptide (secreted and					
CC	transmembrane). The PRO protein, and PRO agonists, PRO antagonists or					
CC	anti-PRO antibodies are useful for preparation of a medicament useful in					

CC the treatment of a condition which is responsive to the PRO protein,  
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be  
CC employed as molecular weight markers for protein electrophoresis. The PRO  
CC coding sequence has applications in molecular biology, including use as  
CC hybridisation probes, and in chromosome and gene mapping.

XX Sequence 230 AA;

Query Match 100.0%; Score 1174; DB 22; Length 230;

Best Local Similarity 100.0%; Pred. No. 4.1e-117;

Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLGLQVGYIIGLLGLLTIVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60

|||||  
Db 1 maslglqlvgyilgllglgtlvamllpswktssyvgasivtavgfskglwmeathstg 60

QY 61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISLACIIISVGMRCCTVFCQESRAKDRVAVA 120

|||||  
Db 61 itqcdiystllglpadlgaqaqammmvtssaislaciisvvgmrcctvfcqesrakdrvava 120

QY 121 GGVFILGLGLGFIPVAMNHLGILRDFYSPLPVDSMKFGEALYLGIISSLSLIAGII 180

|||||  
Db 121 ggvfllglglgfipvawnlhgilrdfysplvpdsmkfeigealygliisslfsliagii 180

QY 181 LCFSCSSQRNSNYDAYQAQPLATRSPPRGQPPKVKSEFNSYSLTGYV 230

|||||  
Db 181 lcfsscqrnsnydayqaqplatrssprgppkvksefnsysltgyv 230

#### RESULT 8

AAB88342

ID AAB88342 standard; Protein; 230 AA.

XX

AC AAB88342;

XX

DT 23-MAY-2001 (first entry)

XX

DE Human membrane or secretory protein clone PSEC0059.

XX

KW Human; secretory protein; membrane protein; vaccine; gene therapy;

KW rheumatoid arthritis; diabetes.

XX

OS Homo sapiens.

XX

PN EP1067182-A2.

XX

PD 10-JAN-2001.

XX

PF 07-JUL-2000; 2000EP-0114090.

XX

PR 08-JUL-1999; 99JP-0194179.

XX

PR 11-JAN-2000; 2000JP-0118775.

XX

PR 02-MAY-2000; 2000JP-0183766.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX

DR WPI; 2001-093989/11.

XX

DR N-PSDB; AAF93769.

XX

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in

PT gene therapy or as candidate target molecules in drug development -

XX

PS Claim 1; SEQ ID 52; 609pp + CD ROM; English.

XX

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916

CC which encode human secretory or membrane proteins represented by

CC AAF88317 - AAF88419. Included in the invention are primers

CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

CC sequences, which can be used in vaccines. The polynucleotide sequences  
CC can be used in gene therapy. The polynucleotide sequences and the  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with inappropriate secretory  
CC protein/membrane protein expression. The nucleic acids and complementary  
CC sequences may also be used as DNA probes in diagnostic assays  
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
CC presence of similar nucleic acid sequences in samples. They may also be  
CC used to study the expression and function of secretory proteins/membrane  
CC polypeptides and their role in metabolism. The polypeptides may be used  
CC as antigens in the production of antibodies against them and in assays to  
CC identify modulators (agonists and antagonists) of expression and  
CC activity. The antibodies and antagonists may also be used as therapeutic  
CC agents to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of the  
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
CC arthritis and diabetes.

XX Sequence 230 AA;

Query Match 100.0%; Score 1174; DB 22; Length 230;

Best Local Similarity 100.0%; Pred. No. 4.1e-117;

Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLGLQVGYIIGLLGLLTIVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60

|||||  
Db 1 maslglqlvgyilgllglgtlvamllpswktssyvgasivtavgfskglwmeathstg 60

QY 61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAISLACIIISVGMRCCTVFCQESRAKDRVAVA 120

|||||  
Db 61 itqcdiystllglpadlgaqaqammmvtssaislaciisvvgmrcctvfcqesrakdrvava 120

QY 121 GGVFILGLGLGFIPVAMNHLGILRDFYSPLPVDSMKFGEALYLGIISSLSLIAGII 180

|||||  
Db 121 ggvfllglglgfipvawnlhgilrdfysplvpdsmkfeigealygliisslfsliagii 180

QY 181 LCFSCSSQRNSNYDAYQAQPLATRSPPRGQPPKVKSEFNSYSLTGYV 230

|||||  
Db 181 lcfsscqrnsnydayqaqplatrssprgppkvksefnsysltgyv 230

#### RESULT 9

AAB66127

ID AAB66127 standard; protein; 230 AA.

XX

AC AAB66127;

XX

DT 02-APR-2001 (first entry)

XX

DE Protein of the invention #39.

XX

KW Secreted; transmembrane; gene therapy.

XX

OS Unidentified.

XX

PN WO200078961-A1.

XX

PD 28-DEC-2000.

XX

PF 18-FEB-2000; 2000WO-US04342.

XX

PR 23-JUN-1999; 99US-0141037.

XX

PR 20-JUL-1999; 99US-0144758.

XX

PR 26-JUL-1999; 99US-0145698.

XX

PR 01-SEP-1999; 99WO-US20111.

XX

PR 29-OCT-1999; 99US-0162506.

XX

PR 30-NOV-1999; 99WO-US28313.

XX

PR 02-DEC-1999; 99WO-US28551.

XX

PR 16-DEC-1999; 99WO-US30095.

XX

PR 05-JAN-2000; 2000WO-US00219.

XX

PR 06-JAN-2000; 2000WO-US00376.

```
XX (GERTH ) GENENTECH INC.
PA Baker RP, Bolstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurrey AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Matanabe CK, Williams PM, Wood WT;
XX WPI: 2001-071395/08.
DR
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX Claim 1: Fig 78; 787pp; English.
PS
XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
XX Sequence 230 AA:
SQ
XX
XX Query Match 100.0%; Score 1174; DB 22; Length 230;
XX Best Local Similarity 100.0%; Pred. No. 4.1e-117;
XX Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MASLGQLVGYILGLLGLTLVAMLLPSWKTSSYVGASITVAVGFSKGLMNECATSTG 60
DB 1 maslglqlvgyilgllgltlvamlpswktssyvgasitvavgfsgklmnecatstg 60
OY 61 ITQCDIYSTILGLPADIQAAQAMWYSSAISLACIISVGMKCTVFCQESAKRARA 120
DB 61 itqcdiystilglpadlqaagammvtsaislaciisvgmctvfcqesakrardrava 120
OY 121 GGVFFILGLGLFIPVAMNLGILRDFYSPVPSMKFEIGALYLGITISLFSLIAGIT 180
DB 121 ggvffilglglfipvamnlgilrdfysplvpdskfeligalylgitslfsliagiti 180
OY 181 LCFSCSSQNRNSNYDAYQAQPLATRSSPPRGQPPKVKSEFNSYSLTGYV 230
DB 181 lcfsscqnrmnsnydayqaqplatrsspprgqppkvksefnsysltgyv 230
XX
XX RESULT 10
XX AAM38857
XX ID AAM38857 standard; Protein: 230 AA.
XX
XX AAM38857;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 2002.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX OS Homo sapiens.
XX
XX PN WO20015312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.
XX
XX
```

```
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX
XX N-PSDB: AAI58013.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 3: SEQ ID NO 2002; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilization of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemia and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 230 AA:
SQ
XX
XX Query Match 99.6%; Score 1169; DB 22; Length 230;
XX Best Local Similarity 99.6%; Pred. No. 1.4e-116;
XX Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 MASLGQLVGYILGLLGLTLVAMLLPSWKTSSYVGASITVAVGFSKGLMNECATSTG 60
DB 1 maslglqlvgyilgllgltlvamlpswktssyvgasitvavgfsgklmnecatstg 60
OY 61 ITQCDIYSTILGLPADIQAAQAMWYSSAISLACIISVGMKCTVFCQESAKRARA 120
DB 61 itqcdiystilglpadlqaagammvtsaislaciisvgmctvfcqesakrardrava 120
OY 121 GGVFFILGLGLFIPVAMNLGILRDFYSPVPSMKFEIGALYLGITISLFSLIAGIT 180
DB 121 ggvffilglglfipvamnlgilrdfysplvpdskfeligalylgitslfsliagiti 180
OY 181 LCFSCSSQNRNSNYDAYQAQPLATRSSPPRGQPPKVKSEFNSYSLTGYV 230
DB 181 lcfsscqnrmnsnydayqaqplatrsspprgqppkvksefnsysltgyv 230
XX
XX RESULT 11
XX AAM40643
XX ID AAM40643 standard; Protein: 260 AA.
XX
XX AAM40643;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5574.
XX
XX
```

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimers; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX PI WPI; 2001-442253/47.

XX DR N-PSDB; AAI59799.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

XX PT Example 2; SEQ ID NO 5574; 10078pp; English.

XX PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

XX CC Note: The sequence data for this patent did not form part of the printed specification.

XX SQ Sequence 260 AA;

Query Match 99.6%; Score 1169; DB 22; Length 260;  
Best Local Similarity 99.6%; Pred. No. 1.7e-116;  
Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MASIGLQVGYILGLGLGLTGLVAMLLPSWKTSSVYGASIVTAVGFSKGLWMECATHTSG 60  
|||||  
Db 31 masigqlqvgyilglglglvamlplswktssvvgasivtavgfskglwmeccathtsg 90  
|||||

Oy 61 ITQCDIYSTLLGLPADIAQAAMVTSSAIISSSLACIIISVGMRCITVFCQESRAKDRVAVA 120  
|||||  
Db 91 itqcdiystllglpadiaqaamvmtssaislaciisvvgmrcitvfcqesrakdrava 150  
|||||

Oy 121 GGVFFILGGLGIFPVAWNLHGILRDPYSPVLPDMSKMFETGEALYLGIISSLSFIAGII 180  
|||||

Db 151 ggvffilgllgfipvawnlhgilrdfyaplvdpmskfeigealylgiisslfsliagii 210

Oy 181 LCFSCSSQNRNRYDAYQAQPLATRSRPPGQPKVKSEFNYSYTGYY 230  
|||||

Db 211 lcfscscqrnsnydaydaqplatrssrppgqpkvksefnysyitgyv 260  
|||||

RESULT 12

AAI36134

ID AAY36134 standard; protein; 230 AA.

XX AC AAY36134;

XX DT 23-SEP-1999 (first entry)

XX DE Human secreted protein #6.

XX KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic; diagnostic; gene therapy; chromosome mapping; secretion vector.

XX OS Homo sapiens.

XX PN WO9925825-A2.

XX PD 27-MAY-1999.

XX PF 13-NOV-1998; 98WO-IB01862.

XX PR 04-SEP-1998; 98US-0099273.

XX PR 13-NOV-1997; 97US-0066677.

XX PR 17-DEC-1997; 97US-0069957.

XX PR 09-FEB-1998; 98US-0074121.

XX PR 13-APR-1998; 98US-0081563.

XX PR 10-AUG-1998; 98US-0096116.

XX PA (GSET ) GENSET.

XX PI Bougueleret L, Duciert A, Dumas Milne Edwards J;

XX DR WPI; 1999-347472/29.

XX DR N-PSDB; AAX97818.

XX PT Extended cDNAs encoding secreted proteins

XX PS Example 28; Page 234-235; 307pp; English.

XX CC AAY36129-Y36222 represent novel human secreted proteins encoded by the extended cDNA sequences represented in AAX97813-X97906. The proteins of the invention have cytostatic, thrombotic and osteopathic activity. The extended cDNAs can be used to express secreted proteins or parts of them or to obtain antibodies capable of binding to the secreted proteins. They may also be used in diagnostic, forensic, gene therapy and chromosome mapping procedures. Uses also include design of expression vectors and secretion vectors.

XX SQ Sequence 230 AA;

Query Match 98.5%; Score 1156; DB 20; Length 230;  
Best Local Similarity 98.7%; Pred. No. 3.4e-115;  
Matches 227; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MASIGLQVGYILGLGLGLTGLVAMLLPSWKTSSVYGASIVTAVGFSKGLWMECATHTSG 60  
|||||

Db 1 masigqlqvgyilglglglvamlplswktssvvgasivtavgfskglwmeccathtsg 60  
|||||

Oy 61 ITQCDIYSTLLGLPADIAQAAMVTSSAIISSSLACIIISVGMRCITVFCQESRAKDRVAVA 120  
|||||

Db 61 itqcdiystllglpadiaqaamvmtssaislaciisvvgmrcitvfcqesrakdrava 120  
|||||

Oy 121 GGVFFILGGLGIFPVAWNLHGILRDPYSPVLPDMSKMFETGEALYLGIISSLSFIAGII 180  
|||||

Db 121 ggvffilgllgfipvawnlhgilrdfyaplvdpmskfeigealylgiisslfsliagii 180  
|||||

OY 181 LCFSCSSQRNRKSNRYDAYQAQPIATRSSPPGPPKVKSEPNSTLTGYV 230  
DB 181 LCFSCSSQRNRKSNRYDAYQAQPIATRSSPPGPPKVKSEPNSTLTGYV 230

## RESULT 13

AAV51676  
ID AAV51676 standard; Protein; 230 AA.

AAV51676;

02-JUN-2000 (first entry)

Murine clodin 2 protein.

Clodin 2; murine; tight junction-constituting membrane protein;  
medicine.

Mus sp.

JP2000032984-A.

02-FEB-2000.

26-JUN-1998; 98JP-0179847.

15-MAY-1998; 98JP-0133215.

(EISA) EISAI CO LTD.

WPI; 2000-285512/25.

DR N-PSDB; AA289137.

Tight junction-constituting membrane protein clodin family - useful in  
the medical field

Claim 2; Page 10; 22pp; Japanese.

This invention describes novel murine nucleic acid sequences encoding the  
clodin family of tight junction (TJ)-constituting membrane protein. The  
CC membrane protein can be used in medical field. This sequence represents  
CC the clodin 2 protein described in the method of the invention.  
CC

Sequence 230 AA;

Query Match 93.4%; Score 1096; DB 21; Length 230;  
Best Local Similarity 91.3%; Pred. No. 8.8e-109;  
Matches 210; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

OY 1 MASTGLIVGYITLGLGLCTLVAMLLPSMKTSSVYGVASTVYANGFSGKGLMECATHTSTG 60  
DB 1 MASTGLIVGYITLGLGLCTLVAMLLPSMKTSSVYGVASTVYANGFSGKGLMECATHTSTG 60  
OY 61 ITQCDDYITLGLPADIOAAQAMVTSATSLACITSVGMRCCTVCOESRAKRDVAVA 120  
DB 61 ITQCDDYITLGLPADIOAAQAMVTSATSLACITSVGMRCCTVCOESRAKRDVAVA 120  
OY 121 GGVFLLGLGLFIPVAMNLHGILRDFYSPLVPDSMKFEIGALYLGITSLFSLIAGII 180  
DB 121 GGVFLLGLGLFIPVAMNLHGILRDFYSPLVPDSMKFEIGALYLGITSLFSLIAGII 180  
OY 181 LCFSCSSQRNRKSNRYDAYQAQPIATRSSPPGPPKVKSEPNSTLTGYV 230  
DB 181 LCFSCSSQRNRKSNRYDAYQAQPIATRSSPPGPPKVKSEPNSTLTGYV 230

## RESULT 14

AAE04228  
ID AAE04228 standard; Protein; 140 AA.

AAE04228;

XX 09-AUG-2001 (first entry)  
XX  
DE Human gene 10 encoded secreted protein HTP1H83, SPO ID NO:83.  
XX

Human; secreted protein; proliferative disorder; cancer; tumour;  
foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
cardiovascular disorder; angiogenic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
endocrine disorder; infection; wound healing; vulneryary;  
cell culture; chemotaxis; food additive;  
binding partner identification.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..24 /label= signal\_peptide

Protein 25..140 /note= "Mature secreted protein"

Misc-difference 136 /note= "Encoded by GNT"

WO200136432-A2.

25-MAY-2001.

15-NOV-2000; 2000WO-US31162.

19-NOV-1999; 99US-0166415.

30-JUN-2000; 2000US-0215136.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsuolis GA, Baker KP, Young PE;

WPI; 2001-343793/36.

DR N-PSDB; AAD08518.

Isolated nucleic acid molecule encoding a human secreted protein is  
used in preventing, treating or ameliorating a medical condition -  
Claim 11; Page 459; 509pp; English.

AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
AAE04240-AAE04297 represent human secreted protein fragments or variants.  
The secreted proteins and their genes are useful for preventing,  
treating or ameliorating medical conditions, e.g., by protein or gene  
therapy. Pathological conditions can be diagnosed by determining the  
amount of the new protein in a sample or by determining the presence of  
mutations in the new genes. Specific uses are described for each of the  
18 genes, based on the tissues in which they are most highly expressed,  
and include developing products for the diagnosis or treatment of  
proliferative disorders, cancer, tumours, foetal and developmental  
abnormalities, haematopoietic disorders, diseases of the immune system,  
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
allergies, neurological disorders (e.g., Alzheimer's disease,  
Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
cardiovascular disorders, angiogenic disorders, kidney disorders,  
gastrointestinal disorders, pregnancy-related disorders, endocrine  
disorders, and infections. The proteins can also be used to aid wound  
healing and epithelial cell proliferation, to prevent skin ageing due to  
sunburn, to maintain organs before transplantation, for supporting cell  
culture of primary tissues, to regenerate tissues, to identify their  
cognate ligands or binding partners, and in chemotaxis, and can be used  
as a food additive or preservative to modify storage properties.  
Antibodies specific for a protein of the invention can be used in

CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.  
XX  
SQ Sequence 140 AA;

Query Match 59.2%; Score 695; DB 22; Length 140;  
Best Local Similarity 98.6%; Pred. No. 3.2e-66;  
Matches 138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MASGLQLVGYILGLLGLTLVAMLLPSWKTSSYVVGASIVTAVGFSKGLWMECATHSTG 60  
Db 1 maslgqlvgyilgllgltlvamlpswktssyvgasivtavgfskglwmeathstg 60  
QY 61 ITQCDIYSTLLGLPADIQAAQAMMTSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120  
Db 61 itqcdiystllglpadlqaaqammtssaislaciisvvgmrctvfcqesrakdrvava 120  
QY 121 GGVEFFILGGLGFIPVAVNL 140  
Db 121 ggveffilgslgflpxawnl 140

RESULT 15  
AAB54052  
ID AAB54052 standard; Protein; 155 AA.  
AC AAB54052;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:504.  
XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative.  
XX

OS Homo sapiens.  
XX  
XX  
PN WO200055320-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05989.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
XX  
DR WPI: 2000-579444/54.  
DR N-PSDB; AAC98817.  
XX

XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
PS Claim 11; Page 942-943; 1379pp; English.  
XX  
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays

CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX

SQ Sequence 155 AA;

Query Match 54.1%; Score 635; DB 21; Length 155;  
Best Local Similarity 97.7%; Pred. No. 9.4e-60;  
Matches 127; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MASLGLQLVGYILGLLGLTLVAMLLPSWKTSSYVVGASIVTAVGFSKGLWMECATHSTG 60  
Db 26 maslgqlvgyilgllgltlvamlpswktssyvgasivtavgfskglwmeathstg 85  
QY 61 ITQCDIYSTLLGLPADIQAAQAMMTSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120  
Db 86 itqcdiystllglpadlqaaqammtssaislaciisvvgmrctvfcqesrakdrvava 145  
QY 121 GGVEFFILGGL 130  
Db 146 ggveffilgxl 155

Search completed: June 2, 2002, 20:10:58  
Job time: 6692 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:00:07 ; Search time 3664.46 Seconds  
(without alignments)  
265.191 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_107  
Perfect score: 72  
Sequence: 1 atggcctcttggcctcca.....ttttgggcacactggttgcc 72

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72	100.0	369	10	T86266
2	72	100.0	494	10	BG385562
3	72	100.0	643	10	BG328625
4	72	100.0	822	10	BG386074
5	72	100.0	945	10	BE513091
6	70.4	97.8	808	10	BG469035
7	60.8	84.4	450	9	AV602537
8	60.8	84.4	517	10	BE480970
9	60.8	84.4	551	9	AV605969
10	60.8	84.4	555	9	AV604013
11	60.8	84.4	569	10	BE667438
12	60.8	83.3	728	10	BG325755
13	60.8	83.3	856	10	BF541534
14	60.8	83.3	886	10	BF032123
15	60.8	83.3	978	10	BG164062
16	56.6	78.6	601	9	AV647382
17	56	77.8	348	9	BB843637

18	56	77.8	371	9	BB873329
19	56	77.8	383	9	BB843776
20	56	77.8	403	9	AW259767
21	56	77.8	422	9	BB848306
22	56	77.8	433	9	BB845010
23	56	77.8	483	9	AI116814
24	56	77.8	490	9	AW107793
25	56	77.8	493	9	AI115927
26	56	77.8	501	9	AW258786
27	56	77.8	520	9	AI746647
28	56	77.8	529	9	AI119108
29	56	77.8	593	9	AI789119
30	56	77.8	619	9	BB653578
31	56	77.8	628	9	AI788398
32	56	77.8	659	9	AI789490
33	56	77.8	673	9	AI746723
34	56	77.8	722	9	AI790813
35	56	77.8	750	9	AW475316
36	56	77.8	2829	11	AK004990
37	54.4	75.6	279	9	BB574507
38	54.4	75.6	458	10	BF789570
39	54.4	75.6	704	10	BI330754
40	54.4	75.6	727	10	BI101652
41	54.4	75.6	757	10	BI143586
42	54.4	75.6	758	10	BI102679
43	54.4	75.6	777	10	BI102100
44	54.4	75.6	846	10	BF789255
45	54.4	75.6	1114	10	BF789836

ALIGNMENTS

RESULT	1
T86266	
LOCUS	T86266 359 bp mRNA linear EST 17-MAR-1995
DEFINITION	Yd84h07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:114973 5' similar to SP:A39484 A39484 ANDROGEN-WITHDRAWAL APOPTOSIS PROTEIN RVPL, PROSTATIC - ;, mRNA sequence.
ACCESSION	T86266
VERSION	T86266.1 GI:714618
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 369) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 1623 High quality sequence stops: 288 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1623 Std Error: 0.00 Seq primer: M13RP1 High quality sequence stop: 288. Location/Qualifiers 1. 369 /organism="Homo sapiens" /db_xref="GDB:470590" /db_xref="taxon:9606" /clone="IMAGE:114973" /clone_lib="Soares fetal liver spleen INFLS"
FEATURES	source

```

/sex="male"
/dev stage="20 week-post conception fetus"
/lab host="rdh10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAAAGATTATTAAGAAGCTCTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."
BASE COUNT      73 a    106 c    100 g     87 t     3 others
ORIGIN

```

	Query March	100.0%	Score 72	DB 10	Length 369
	Best Local Similarity	100.0%	Pred. No. 4.2e-12		
	Matches 72	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	1 atggcctctcttggcgctcaactfttgggctacacccagagccttcttgggcttttggc 60				
Db	77 atggcctctcttggcgctcaactfttgggctacacccagagccttcttgggcttttggc 136				
Oy	61 aacactgattgcc 72				
Db	137 acactgattgcc 148				

RESULT	2
LOCUS	BG385562
DEFINITION	BG385562 494 bp mRNA linear EST 12-MAR-2001 602453858e1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4582214 5', mRNA sequence.
ACCESSION	BG385562
VERSION	BG385562.1 GI:13278350
KEYWORDS	EST.
SOURCE	human. Homo sapiens
ORGANISM	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 494) NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	
JOURNAL	
COMMENT	

FEATURES	Location/Qualifiers
source	1. .494

```

/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB1; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT
91 a 145 c 132 g 126 t

```

ORIGIN		Query Match	100.0%	Score 72:	DB 10:	Length 494:
		Best Local Similarity	100.0%	Pred. No. 4.6e-12:		
	Matches	72:	Conservative	0:	Mismatches	0:
					Indels	0:
					Gaps	0:
0y	1 atygacctctctggagcctcoactgtg9gctacatcctagagcctctg9g9cctttt9gac	60				
Db	35 atggcctctcttgggctctccacttg6ggctacatcttagggccttctggggcttttgggc	94				
0y	61 acaactgtgtgac	72				
Db	95 acactggttgcc	106				

RESULT	3				
LOCUS	BG328625				
DEFINITION	643 bp mRNA linear EST 27-FEB-2001				
ACCESSION	602427889p1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4547230 5',				
VERSION	BG328625				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 643)				
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

1253411:readme.txt: file  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.lnl.gov  
Plate: LNCM1235 row: a column: 23  
High quality sequence stopt: 641.  
Location/Qualifiers  
1..643  
FEATURES  
source

```

FEATURES
SOURCE
Location/Qualifiers
1. 643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:454/230"
/clone_id="NH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; site_1: XhoI; site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT
130 a 173 c 177 g 163 t
ORIGIN

```

Query Match	Similarity	Score	72	DB	10	Length	643
Best Local	Similarity	100.0%	Pred.	NO.	4.9e-12		
Matches	72	Conservative	0	Mismatches	0	Indels	0
				Gaps			0
QY	1	atggcctctctgtgctccaactlgtggctacatcctagcctcttggc	60				
Db	116	atggcctctctgtgctccaactlgtggctacatcctagcctcttggc	175				
QY	61	acactggttgc	72				

```

Db 176 ACACGTGGTTGCC 187

RESULT 4
BG386074
LOCUS 602455248F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583675 5',
DEFINITION mRNA sequence.
ACCESSION BG386074
VERSION 1 GI:13279520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1307 row: p column: 12
High quality sequence stop: 746.
FEATURES
source
location/Qualifiers
1..822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583675"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 181 a 224 c 217 g 200 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 822;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctctgtgggctttgggc 60
|||||
Db 127 ATGGCCCTCTTGGCTCCAACTTGTGGCTACATCCTAGGCTTCTGGGCTTTGGGC 186
|||||

QY 61 acactggttggc 72
|||||
Db 187 ACACGTGGTTGCC 198

RESULT 5
BE513091
LOCUS 601171545F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:3545171 5',
DEFINITION mRNA sequence.
ACCESSION BE513091
VERSION 1 GI:9720302
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 945)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM241 row: i column: 12
High quality sequence stop: 782.
FEATURES
source
location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3545171"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 194 a 262 c 263 g 226 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 945;
Best Local Similarity 100.0%; Pred. No. 5.5e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctctgtgggctttgggc 60
|||||
Db 187 ATGGCCCTCTTGGCTCCAACTTGTGGCTACATCCTAGGCTTCTGGGCTTTGGGC 246
|||||

QY 61 acactggttggc 72
|||||
Db 247 ACACGTGGTTGCC 258

RESULT 6
BG469035
LOCUS 602510779F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645075 5',
DEFINITION mRNA sequence.
ACCESSION BG469035
VERSION 1 GI:13401220
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 945)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM241 row: i column: 12
High quality sequence stop: 36
High quality sequence stop: 782.
FEATURES
source
location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3545171"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 194 a 262 c 263 g 226 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 945;
Best Local Similarity 100.0%; Pred. No. 5.5e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctctgtgggctttgggc 60
|||||
Db 187 ATGGCCCTCTTGGCTCCAACTTGTGGCTACATCCTAGGCTTCTGGGCTTTGGGC 246
|||||

QY 61 acactggttggc 72
|||||
Db 247 ACACGTGGTTGCC 258

RESULT 6
BG469035
LOCUS 602510779F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645075 5',
DEFINITION mRNA sequence.
ACCESSION BG469035
VERSION 1 GI:13401220
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLCM417 row: n column: 20  
High quality sequence stop: 585.  
Location/Qualifiers

## FEATURES

source

```
1. 808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
```

BASE COUNT

204 a 216 c 200 g 188 t

ORIGIN

Query Match 97.8%; Score 70.4; DB 10; Length 808;  
Best Local Similarity 98.6%; Pred. No. 1.6e-11;  
Matches 71; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atggcctcttggccctcaactgtgtggtacatcctcaggccttctggccttgggc 60  
|||||  
Db 124 ATGGACCTCTTGCCCTCCACACTGTGGCTACATCTCTAGCCTTCTGGCCTTTTGGGC 183  
|||||

OY 61 acactgtgtgcc 72  
|||||  
Db 184 ACACGTGGTGCC 195  
|||||

## RESULT 7

AV602537 450 bp mRNA linear EST 27-NOV-2001  
LOCUS AV602537 Bos taurus kidney fetus Bos taurus cDNA clone E1K1010H03  
DEFINITION 5' mRNA sequence.

ACCESSION AV602537  
VERSION AV602537.1 GI:9724863  
KEYWORDS EST.

SOURCE

ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 450)  
Takasuga,A., Hirotsune,S., Itoh,R., Jitchozono,A., Suzuki,H., Aso,H.  
and Sugimoto,Y.  
Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)

JOURNAL MEDLINE COMMENT

Contact: Yoshikazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725  
Email: kazusugi@cocoa.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.  
Location/Qualifiers

## FEATURES

source

```
1. 450
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="E1K1010H03"
/clone_lib="Bos taurus kidney fetus"
```

```
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/Note="Vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
```

BASE COUNT

91 a 131 c 135 g 92 t 1 others

ORIGIN

Query Match 84.4%; Score 60.8; DB 9; Length 450;  
Best Local Similarity 90.3%; Pred. No. 1.1e-08;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atggcctcttggccctcaactgtgtggtacatcctcaggccttctggccttgggc 60  
|||||  
Db 136 ATGGCCTCTTGCCCTCCACACTGTGGCTACATCTCTAGCCTTCTGGCCTTTTGGGC 195  
|||||

OY 61 acactgtgtgcc 72  
|||||  
Db 196 ACCGTGATTGCC 207  
|||||

## RESULT 8

BE480970 517 bp mRNA linear EST 28-AUG-2000  
LOCUS BE480970 BARC SBOV Bos taurus cDNA 5' mRNA sequence.

DEFINITION BE480970  
ACCESSION BE480970  
VERSION BE480970.1 GI:9600503  
KEYWORDS EST.

SOURCE

ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 517)  
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and  
Wells,K.D.  
Mapping of Expressed Sequence Tags from a normalized bovine mammary  
gland cDNA library  
Unpublished (2000)  
Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414  
Email: tads@psi.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCATGACCAT  
BACKWARD: GTTTCACGACGACG  
Plate: 18 row: I column: 20  
Seq primer: ATTAGTGACACTATAG.  
Location/Qualifiers

JOURNAL COMMENT

FEATURES

source

```
1. 517
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC SBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
```

BASE COUNT 108 a 144 c 158 g 107 t

ORIGIN

Query Match 84.4%; Score 60.8; DB 10; Length 517;  
Best Local Similarity 90.3%; Pred. No. 1.2e-08;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 1 atggcctctcttgccctccaaactgtgggtacatcctagcctcttggtggttttgggc 60
Db 150 ATGGCCTCTCTTGGCTCCACCTTGTGGGCTACGCTCTGGGCTTCTGGGCTGTGGGC 209

QY 61 acactgggtgccc 72
Db 210 ACCGTGATTGCC 221

RESULT 9
AV605969
LOCUS
DEFINITION AV605969 Bos taurus kidney fetus Bos taurus cDNA clone ELK1032A05
5', mRNA sequence.
ACCESSION AV605969
VERSION AV605969.1 GI:9736342
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source
location/Qualifiers
1..551
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="ELK1032A05"
/clone_lib="Bos taurus kidney fetus"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/notes="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 105 a 151 c 163 g 120 t 12 others
ORIGIN

Query Match 84.4%; Score 60.8; DB 9; Length 551;
Best Local Similarity 90.3%; Pred. No. 1.2e-08;
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctccaaactgtgggtacatcctagcctcttggtggttttgggc 60
Db 141 ATGGCCTCTCTTGGCTCCACCTTGTGGGCTACGCTCTGGGCTTCTGGGCTGTGGGC 200

QY 61 acactgggtgccc 72
Db 201 ACCGTGATTGCC 212

RESULT 10
AV604013
LOCUS
DEFINITION AV604013 Bos taurus kidney fetus Bos taurus cDNA clone ELK1020B11
5', mRNA sequence.
ACCESSION AV604013

```

```

VERSION AV604013.1 GI:9734386
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source
location/Qualifiers
1..555
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="ELK1020B11"
/clone_lib="Bos taurus kidney fetus"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/notes="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 106 a 158 c 167 g 124 t
ORIGIN

Query Match 84.4%; Score 60.8; DB 9; Length 555;
Best Local Similarity 90.3%; Pred. No. 1.2e-08;
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctccaaactgtgggtacatcctagcctcttggtggttttgggc 60
Db 137 ATGGCCTCTCTTGGCTCCACCTTGTGGGCTACGCTCTGGGCTTCTGGGCTGTGGGC 196

QY 61 acactgggtgccc 72
Db 197 ACCGTGATTGCC 208

RESULT 11
BE667438
LOCUS
DEFINITION BE667438 569 bp mRNA linear EST 25-APR-2001
ACCESSION BE667438
VERSION BE667438.1 GI:10028029
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)

```

MEDLINE 21180013  
CONTACT: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGCAGCAGC  
Plate: 64 row: M column: 1  
Seq primer: ATTAGTGAACACTATAG.  
Location/Qualifiers  
1. 569  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 4BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORF6; Site:1: XbaI; Site:2: XhoI;  
library made from pooled tissue from day 20 and day 40  
embryos."  
BASE COUNT 106 a 164 c 168 g 131 t  
ORIGIN  
Query Match 84.4%; Score 60.8; DB 10; Length 569;  
Best Local Similarity 90.3%; Pred. No. 1.2e-08;  
Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 atggcctctcttgagcccaacttggtggtacatcctagcctcttgaggctttgggc 60  
|||||  
DB 108 ATGGCCTCTTGTGGCCCTCCAGCTTGCTGAGCTACCTCTGCGCCCTTCTGGGCTTTGGGC 167  
|||||  
QY 61 acactggttggc 72  
|||  
DB 168 ACCGTGATGGCC 179  
|||  
RESULT 12  
BG325755 728 bp mRNA linear EST 27-FEB-2001  
LOCUS 60242446P1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4562273 5',  
DEFINITION mRNA sequence.  
ACCESSION BG325755  
VERSION BG325755.1 GI:13132179  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 728)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1CM1274 row: d column: 18  
High quality sequence stop: 708.  
Location/Qualifiers  
1. 728  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="IMAGE:4562273"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 154 a 205 c 183 g 186 t  
ORIGIN  
Query Match 83.3%; Score 60; DB 10; Length 728;  
Best Local Similarity 98.6%; Pred. No. 2.3e-08;  
Matches 71; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 atggcctctcttgagcccaacttggtggtacatcctagcctcttgaggctttgggc 60  
|||||  
DB 30 ATGGCCTCTTGTGGCCCTCCAGCTTGCTGAGCTACCTCTGCGCCCTTCTGGGCTTTGGGC 88  
|||||  
QY 61 acactggttggc 72  
|||||  
DB 89 ACACGTGTTGCC 100  
|||||  
RESULT 13  
BF541534 856 bp mRNA linear EST 11-DEC-2000  
LOCUS 60206790P1 NIH\_MGC\_58 Homo sapiens cDNA clone IMAGE:4067065 5',  
DEFINITION mRNA sequence.  
ACCESSION BF541534  
VERSION BF541534.1 GI:11628915  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 856)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1CM905 row: k column: 02  
High quality sequence stop: 597.  
Location/Qualifiers  
1. 856  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4067065"  
/clone\_lib="NIH\_MGC\_58"  
/tissue\_type="hypernephroma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site:1:  
SfiI (ggcgccctcgcc); Site:2: SfiI (ggccatctagcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTAAGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGGCGGCGGACATG-dT(30)BN-3'  
(where B = A, C, G or N = A, C, G, or T). Average  
insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)." 181 t 1 others

BASE COUNT 200 a 222 c 252 g 181 t 1 others

ORIGIN

Query Match 83.3%; Score 60; DB 10; Length 856;  
Best Local Similarity 98.6%; Pred. No. 2.4e-08;  
Matches 71; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctcttgggcttttgggc 60  
|||||  
Db 138 ATGGCCCTCTCTGGCTCCAACTTGTGGGTACATCCTAGGCTTCTGGGC 196  
|||||

QY 61 acactggttgc 72  
|||||

Db 197 ACACGTGTGCC 208

RESULT 14  
BF032123 886 bp mRNA linear EST 10-OCT-2000  
LOCUS 601559727F1 NIH\_MGC\_58 Homo sapiens cDNA clone IMAGE:3829254 5',  
DEFINITION mRNA sequence.  
ACCESSION BF032123  
VERSION BF032123.1 GI:10739835  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 886)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM501 row: f column: 07  
High quality sequence stop: 584.  
Location/Qualifiers  
1..886  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3829254"  
/clone\_lib="NIH\_MGC\_58"  
/tissue\_type="hypernephroma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgcttcggcc); Site\_2: SfiI (ggcattatggcc);  
Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGGCGGCACATG-dT(30)BN-3',  
(Where B = A, C, G, or N = A, C, G, or T). Average  
insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 199 a 219 c 262 g 206 t

ORIGIN

Query Match 83.3%; Score 60; DB 10; Length 886;  
Best Local Similarity 98.6%; Pred. No. 2.4e-08;  
Matches 71; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctcttgggcttttgggc 60  
|||||  
Db 128 ATGGCCCTCTCTGGCTCCAACTTGTGGGTACATCCTAGGCTTCTGGGC 186  
|||||

QY 61 acactggttgc 72  
|||||

Db 187 ACACGTGTGCC 198

RESULT 15  
BG164062 978 bp mRNA linear EST 06-FEB-2001  
LOCUS 602341087F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4449212 5',  
DEFINITION mRNA sequence.  
ACCESSION BG164062  
VERSION BG164062.1 GI:12670765  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 978)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10232 row: m column: 21  
High quality sequence stop: 689.  
Location/Qualifiers  
1..978  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4449212"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypernephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 220 a 279 c 276 g 203 t

ORIGIN

Query Match 83.3%; Score 60; DB 10; Length 978;  
Best Local Similarity 98.6%; Pred. No. 2.5e-08;  
Matches 71; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atggcctcttggcctccaaactgtgggtacatcctagcctcttgggcttttgggc 60  
|||||  
Db 25 ATGGCCCTCTCTGGCTCCAACTTGTGGGTACATCCTAGGCTTCTGGGC 83  
|||||

QY 61 acactggttgc 72  
|||||

Db 84 ACACGTGTGCC 95

Search completed: June 2, 2002, 20:00:09  
Job time: 14619 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:57:16 ; Search time 2378.87 Seconds  
(without alignments)  
6069.819 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_725  
Perfect score: 690  
Sequence: 1 atggcctctcttggcctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

```

Minimum DB'seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*
```

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

## ALIGNMENTS

```

RESULT      1
AF250558
LOCUS      693 bp mRNA linear PRI 09-AUG-2000
DEFINITION Homo sapiens claudin-2 mRNA, complete cds.
ACCESSION AF250558
VERSION AF250558.1 GI:9755008
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 693)
AUTHORS   Reinecker,H.-C., Sakaguchi,T. and Golden,H.M.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAR-2000) Gastrointestinal Unit, Massachusetts
            General Hospital, Fruit Street, Boston, MA 02114, USA
FEATURES   Location/Qualifiers
            1..693
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="Xq22.3-q23"
            /cell_line="T-84"

```

CDs

1 693  
/tissue\_type="epithelium"  
/note="light junction-associated protein"  
/codon\_start=1  
/product="claudin-2"  
/protein\_id="AAF98151.1"  
/db\_xref="GI:9755009"  
/translation="MASLQLQVYILGLGLGVMLPSPKTSYVASIVAV  
GTSKGLMECAVHSTGTIOCDIYSLGLPADIQAAMVWTSLSAISLACISVGM  
KCTVPCQESRANRVAVAGVGFILGLGFLPVANNLHILDFYSPVLPDSMKFEI  
GSLYLGITLSLFLSLIAGIILFCSSSRNSRNYDAYOAPLATRSSPRGPQPKVK  
SEFNSYSLGYV"

BASE COUNT 142 a 202 c 168 g 181 t  
ORIGIN

Query Match 100.0%; Score 690; DB 9; Length 693;  
Best Local Similarity 100.0%; Pred. No. 1.4e-191;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctcttggcctccaactgttggtacatcttagcctcttgggcttgggc 60  
DB 1 ATGGCCTCTTGGCCTCCAACTTGTGGCTACATCTAGGCTTGGGGCTTTGGGC 60  
OY 61 aaactgttgccatgctgctcccaagctggaacaagcttctatgtcgggtccagcatt 120  
DB 61 AACCTGTTGGCCATGCTGCCAGCTGGAACCAAGTCTTATGTCGGTCCAGCATTT 120  
OY 121 gtacagcaattgctctctccaaggcctctggaatggaatgtgcacacacagcagc 180  
DB 121 GTACAGCAATTTGGCTTCTCCAAAGGCTCTGATGGAATGTGCCACACACAGCAGCAG 180  
OY 181 atcaccagctgtgacatctatagaccctctgggctgcccgtgacatcccaagctgccc 240  
DB 181 ATCACCAGTGTGACATCTATAGACACCTTCTGGGCTGCCCTGACATCCAGCTGCC 240  
OY 241 caggcctatgtgtgacatccagctgcaatctcctccctgagctgcatctctgtgtg 300  
DB 241 CAGGCCTATGTGTGACATCTGCAATCTCCCTGGCTGCAATCTCTGTGTG 300  
OY 301 ggcatagatagcacatctcttgcagaagatcccgagacaaagacagatggcgtagca 360  
DB 301 GGCATAGATGCAACATCTCTGCCAGGAATCCGAGCCAAAGACAGATGGCGGTAGCA 360  
OY 361 ggttgagctcttctcatccttggagcctcctggatcttctgttggcttggatctt 420  
DB 361 GGTGAGCTCTTTTCAATCCTTTGGAGGCTCTCTGGATTTCTCTGTTCCTGGCAATCTT 420  
OY 421 catggatctctacaggaacttctacccaactggtgcctgacagcatgaaatttgagatt 480  
DB 421 CATGGATCTCTACGGGACTTCTACTCACCTGGTGCCTGACAGCATGAATTTGAGATT 480  
OY 481 ggaagagcttctacttggcattattcttccctgttccctggatagctggaattatc 540  
DB 481 GGAGAGGCTCTTACTTGGCATTTATTTCTCTGTTCTCCGTGATGCGGGAATATC 540  
OY 541 ctctgcttctctgctcatcccgagaatcgtctccaactactacgaatgcttaccagcc 600  
DB 541 CTCTGCTTTTCTGCTCATCTCCAGAGAAATCGCTCAACTGATGCTTACCAAGCC 600  
OY 601 caacctcttgccacaagagctctccaagcctgtgtcaacctcccaagtcaagaatgag 660  
DB 601 CAACCTCTTGCCACAAGAGCTCTCCAAAGCCCTGGTCAACCTCCCAAGTCAAGATGAG 660  
OY 661 ttcaattctacagcctgacaggtatgtg 690  
DB 661 TTCATTTCTTACAGCCTGACAGGGATATGTG 690

RESULT 2  
AX092348 1475 bp DNA linear PAT 21-MAR-2001  
LOCUS AX092348  
DEFINITION Sequence 79 from Patent WO0116318.

ACCESSION AX092348  
VERSION AX092348.1 GI:134444488  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1475)  
AUTHORS Eaton,D.L., Filvaroff,E., Gerltsen,M.E., Goddard,A.,  
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
Wood,W.I.  
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding  
JOURNAL the same  
PATENT Patent: WO 0116318-A 79 08-MAR-2001;  
Genentech, Inc. (US)  
FEATURES  
source location/Qualifiers  
1..1475  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 345 a 404 c 376 g 350 t  
ORIGIN

Query Match 100.0%; Score 690; DB 6; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 1.3e-191;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctcttggcctccaactgttggtacatcttagcctcttgggcttgggc 60  
DB 122 ATGGCCTCTTGGCCTCCAACTTGTGGCTACATCTAGGCTTGGGGCTTTGGGC 181  
OY 61 aaactgttgccatgctgctcccaagctggaacaagcttctatgtcgggtccagcatt 120  
DB 182 AACCTGTTGGCCATGCTGCCAGCTGGAACCAAGTCTTATGTCGGTCCAGCATTT 241  
OY 121 gtacagcaattgctctctccaaggcctctggaatggaatgtgcacacacagcagc 180  
DB 242 GTACAGCAATTTGGCTTCTCCAAAGGCTCTGGAATGTGCCACACACAGCAGCAG 301  
OY 181 atcaccagctgtgacatctatagaccctctgggctgcccgtgacatcccaagctgccc 240  
DB 302 ATCACCAGTGTGACATCTATAGACACCTTCTGGGCTGCCCTGACATCCAGCTGCC 361  
OY 241 caggcctatgtgtgacatccagctgcaatctcctccctgagctgcatctctgtgtg 300  
DB 362 CAGGCCTATGTGTGACATCTGCAATCTCTGGCTGCAATCTCTGTGTG 421  
OY 301 ggcatagatagcacatctcttgcagaagatcccgagacaaagacagatggcgtagca 360  
DB 422 GGCATAGATGCAACATCTCTGCCAGGAATCCGAGCCAAAGACAGATGGCGGTAGCA 481  
OY 361 ggttgagctcttctcatccttggagcctcctggatcttctgttggcttggatctt 420  
DB 482 GGTGAGCTCTTTTCAATCCTTTGGAGGCTCTCTGGATTTCTCTGTTCCTGGAAATCTT 541  
OY 421 catggatctctacaggaacttctacccaactggtgcctgacagcatgaaatttgagatt 480  
DB 542 CATGGATCTCTACGGGACTTCTACTCACCTGGTGCCTGACAGCATGAATTTGAGATT 601  
OY 481 ggaagagcttctacttggcattattcttccctgttccctggatagctggaattatc 540  
DB 602 GGAGAGGCTCTTACTTGGCATTTATTTCTCTGTTCTCCGTGATGCGGGAATATC 661  
OY 541 ctctgcttctctgctcatcccgagaatcgtctccaactactacgaatgcttaccagcc 600  
DB 662 CTCTGCTTTTCTGCTCATCTCCAGAGAAATCGCTCAACTGATGCTTACCAAGCC 721  
OY 601 caacctcttgccacaagagctctccaagcctgtgtcaacctcccaagtcaagaatgag 660  
DB 722 CAACCTCTTGCCACAAGAGCTCTCCAAAGCCCTGGTCAACCTCCCAAGTCAAGATGAG 781  
OY 661 ttcaattctacagcctgacaggtatgtg 690  
DB 661 1475 bp DNA linear PAT 21-MAR-2001

```
Db 782 TTCAATTCCTACAGCCTGACAGGCTATGTG 811
RESULT 3
LOCUS AX299996 1475 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 1 from Patent W00166740.
ACCESSION AX299996
VERSION AX299996.1 GI:17129473
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Eaton,D.L., Fong,S., Goddard,A., Godowski,P.J., Grimaldi,C.J.,
Gurney,A.L., Tumas,D., Metanabe,C.K., Wood,W.I. and Zhang,Z.
TITLE Compositions and methods for the treatment of immune related
diseases
JOURNAL Patent: WO 0166740-A 1 13-SEP-2001;
Genentech, Inc. (US)
FEATURES
source
1..1475
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 345 a 404 c 376 g 350 t
ORIGIN
Query Match 100.0%; Score 690; DB 6; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1.5e-191;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggcctctcttgccctcccaactgttggtctacatcctagcctctctgggcttttgggc 60
|||||
Db 122 ATGGCTCTCTTGGCTCCAACTTGTGGCTACATCTAGGCTTCTGGGCTTTTGGGC 181
QY 61 acactggttgccatgtctgccctccagctggaacaaagtcttatgtctggtgcccagcatt 120
|||||
Db 182 ACACCTGGTTGCCATGCTGCTCCCGCAGCTGGAAACAAAGTTCTTATGTGCGTGGCCAGCAT 241
QY 121 gtacagcagttggtctctccagggtcctctggtgaatgtgacacacacagcagc 180
|||||
Db 242 GTACAGCAGCTGGCTTCTCCAAAGGGCTCTGGATGGAATGTGCACACAGCAGCAGGC 301
QY 181 atcaccagtgatcatctatagaccctctgggctgcccctgacctgacatccagctgcc 240
|||||
Db 302 ATACCCAGTGTGACATCTATAGCACCTTCTGGGCTGCCCCGCTGACATCCAGGCTGCC 361
QY 241 caggccatgatggtgacatccagtgcaatctctccctggcctgcattctctgtggtg 300
|||||
Db 362 CAGGCCATGATGTTGACATCCAGTGCAATCTCTCCCTGGCTGCATATCTCTGTGTG 421
QY 301 ggcagatgacagctctctgccaggaatcccgagccaaagacagagtgccggtagca 360
|||||
Db 422 GGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 481
QY 361 ggtgagatcttttcccttgaggcctcctggattcattctgttgcctggaattctt 420
|||||
Db 482 GGTGGAGATCTTTTCACTGTGGAGGCTCTCTGGGATTCATCTCTGTGTGCTGGAATCTT 541
QY 421 catggatctctacgggacttctactaccactggtgctgacagcatgaaatttgagatt 480
|||||
Db 542 CATGGGATCTCTACGGGACTTCTACTCACCACTGCTGCTGACACATGAATTTGAGATT 601
QY 481 ggagagcctcttacttgggcattattctccctgttctccctgatagctgggaatc 540
|||||
Db 602 GGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTCTCCCTGATAGCTGGAATCATC 661
QY 541 cctgtcttttctgctcatcccgagagaaatcgctccaactactacgtacgtaccagcc 600
|||||
Db 662 CTCTGCTTTTCTGCTCATCCCGAGAGAAATCGCTCCAACTACTACGATGCTTACCAAGCC 721
```

```
QY 601 caacctcttgccacaggagctctccaaagcctggtcaacctccccaaagtcaagagtga 660
|||||
Db 722 CAACCTCTTTGCCACAAGGAGCTCTCCAAGGCTGGTCAACCTCCCAAGTCAAGAGTGAG 781
QY 661 ttcaattctacagcctgacaggttatgtg 690
|||||
Db 782 TTCAATTCCTACAGCCTGACAGGCTATGTG 811
RESULT 4
LOCUS BC014424 1506 bp mRNA linear PRI 19-SEP-2001
DEFINITION Homo sapiens, Similar to claudin 2, clone MGC:20191 IMAGE:4645075,
mRNA, complete cds.
ACCESSION BC014424
VERSION BC014424.1 GI:15680158
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1506)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Nees, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL plate: 29 Row: m Column: 11.
```

Location/Qualifiers

source

1..1506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MGC:20191 IMAGE:4645075"

/tissue\_type="Colon, adenocarcinoma"

/clone\_lib="NIH MGC\_15"

/lab\_host="DH10B-R"

/note="Vector: pORF7"

122..814

/codon\_start=1

/product="Similar to claudin 2"

/protein\_id="AAHL4424.1"

/db\_xref="GI:15680159"

/translation="MASGLQLVGLGLGLGLVLAFLPSWTSYVGVASIVTAV

GFSGKLMNECATHTGTCQDIYTLGLPADIAQAAMVTSSAISLACIISVYGM

RCTVFCQSRADRVAVAGGVFFILGLGTFPVANLHGLDRFYSPLVPSMKEFI

GEALYLGIISSLSIAGIILCFSCSSQRNRSNYDAYQAQPLATRSSPRQPPKVK

SEFNSYSLTGYV"

BASE COUNT 377 a 404 c 376 g 349 t

ORIGIN

Query Match 100.0%; Score 690; DB 9; Length 1506;

Best Local Similarity 100.0%; Pred. No. 1.5e-191;

Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttgccctcaacttggtggtacatccatagcctctcgggcttggc 60  
|||||  
Db 122 ATGGCCTCTCTTGGCTCCAACTTGTGGCTACATCCTAGGCTTCTGGGCTTTGGGC 181  
Qy 61 aacctggttgccatctgctcccccagctggaataaagaattcttatgtcggtgcagcatt 120  
|||||  
Db 182 ACACGTGTTCACATGCTGCTCCAGCTGGAATAAACAAGTTCTATGTCGGTCCAGCAT 241  
Qy 121 gtgacagcattggtgcttctccaaaggcctctgtgaatgagtcacacacagacaggc 180  
|||||  
Db 242 GTACACAGCATTTGGCTTCTCCAGGCTCTGATGGAATGTGCCACACAGCAGCAGC 301  
Qy 181 atcaccagctgtacatctatagcaacctcttggtgctgcccgtgacatccaaagctgac 240  
|||||  
Db 302 ATCACCAGGTGACATCTATAGCACCTCTGCGGCTGCGCGGTGACATCCAGGCTGCC 361  
Qy 241 caagcagatggtgtacatccagtgcaatctcctcctggtgctgcatctctgtgtg 300  
|||||  
Db 362 CAGGCCATGATGATGACATCAGTGCATCTCCCTCCGCTGCTCATATCTCTGTGGTG 421  
Qy 301 ggcataagatgacaaagtctctccaaagaaatcccgagccaaagaagaagtggtgagca 360  
|||||  
Db 422 GGCATAGATGACAGATCTTCTGCGAGAAATCCGAGCCAAAGACAGAGTGGCGGTAGCA 481  
Qy 361 ggtgagctcttcttccatctcttgagagcctcctggtgattctcctgtgctggaatctt 420  
|||||  
Db 482 GGTGAGGTCTTTTTCATCTTGGAGGCTCTGCGGATTCATCTCTGCTGTGCAATCTT 541  
Qy 421 catggaatctcagaggaatctctacacacactggtgctcctcagacatgaaatttgagatt 480  
|||||  
Db 542 CAGGAGATCTTACAGGAGCTTCTACTACACACTGTGCTGCTGACACATGAATAATTGAGATT 601  
Qy 481 ggaagagctcttaactggtgacattattctcctctgctcctcctgagctgagatctc 540  
|||||  
Db 602 GGAAGAGCTTTTACTTGGCATTTATTTCTCTCTGTTCTCCGATGACGAGATCATC 661  
Qy 541 ctctgcttctctgctcatccagagaatcgcctcaactactactagtgctcacaagcc 600  
|||||  
Db 662 CTCTGCTTTTCCGCTATCTCCAGAGAAATCGCTCCAACTACTAGATGCTTACCAAGCC 721  
Qy 601 caactctctgcaacaagagctctcacaagcctggtgcaactcctccaaagtcagaagtgag 660  
|||||  
Db 722 CAACCTCTTCCACAAAGAGCTCTCCAAAGGCTGTGCTCACTCCCAAGTCAAGAGTGAG 781  
Qy 661 ttcaattctcagagcctgagaggtatgtg 690  
|||||  
Db 782 TTCAATTCCTACAGCTCGACAGGATGTG 811

RESULT 5  
AF177340 1918 bp mRNA linear PRI 03-OCT-2000  
DEFINITION Homo sapiens clone SP82 claudin 2 mRNA, complete cds.  
ACCESSION AF177340  
VERSION AF177340.1 GI:10503979  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
Yu,J. and Han,L.H.  
TITLE Novel human cDNA clone with function of inhibiting cancer cell  
growth  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1918)  
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,

Yu,J. and Han,L.H.

TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1999) National Laboratory for Oncogenes & Related  
Genes, Shanghai Cancer Institute, 25/Jm 2200 Xie Tu Road, Shanghai  
200032, P.R. ChinaFEATURES  
source location/qualifiers1..1918  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="SP82"  
520..1212  
/note="similar to Mus musculus claudin 2"

CDS

/codon\_start=1  
/product="Claudin 2"  
/protein\_id="AA017984.1"  
/db\_xref="GI:10503980"  
/translation="MASIGLVNYTIGLIGLITVAMLLPSMKTSSYGVATVAV  
GFSGLMEKCTHSTGTQCDIYITLGLPDIQAOAMNTSSATISLACTISVGM  
RCTVFCESRAKDRVAAGVFTIGLGLFIPVAMNHLGLRDFSPVDSMKEI  
GEALYLDIISLFSLIAGIILCFSCSSQNRNSNYDAYOAPLATRSSPRGPPKVK  
SEFNSYSITGVY"BASE COUNT 486 a 525 c 473 g 434 t  
ORIGIN

Query Match 100.0%; Score 690; DB 9; Length 1918;

Best Local Similarity 100.0%; Pred. No. 1.5e-191;

Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctcttgccctcaacttggtggtacatccatagcctctcgggcttggc 60  
|||||  
Db 520 ATGGCCTCTCTTGGCTCCAACTTGTGGCTACATCCTAGGCTTCTGGGCTTTGGGC 579  
Qy 61 aacctggttgccatctgctcccccagctggaataaagaattcttatgtcggtgcagcatt 120  
|||||  
Db 580 ACACGTGTTCACATGCTGCTCCAGCTGGAATAAAGTTCTATGTCGGTCCAGCAT 639  
Qy 121 gtgacagagctgtgcttctccaaaggcctcctggtgattctcctgtgctcacaagcaagc 180  
|||||  
Db 640 GTGACAGCATTTGGCTTCTCCAAAGGCTCTGAGATGATGTGCCACACAGCAGCAGC 699  
Qy 181 atcaccagctgtgacatctatagcaacctcttggtgctgcccgtgacatccagctgctc 240  
|||||  
Db 700 ATCACCAGGTGACATCTATAGCACCTTCTGCGCTGCTCCGCTGACATCCAGGCTGCC 759  
Qy 241 caggcagatggtgacatccagtgcaatctcctcctggtgattctcctgtgctggaatctt 300  
|||||  
Db 760 CAGGCCATGATGATGACATCAGTGCATCTCCCTGCTGCTGATATCTCTGTGGTG 819  
Qy 301 ggcataagatgacaaagtctctccaaagaaatcccgagccaaagaagaagtggtgagca 360  
|||||  
Db 820 GGCATAGATGACAGATCTTCTGCGAGAAATCCGAGCCAAAGACAGATGCGGTAGCA 879  
Qy 361 ggtgagctcttcttccatctcttgagagcctcctggtgattctcctgtgctggaatctt 420  
|||||  
Db 880 GGTGAGGTCTTTTACTTGGAGGCTCTCCGAGATTCATCTCTGCTCTGGAATCTT 939  
Qy 421 catggaatctcagaggaatctcaccacacagctggtgctgcaagacatgaaatttgagatt 480  
|||||  
Db 940 CATGGATCTTACAGGAGCTTCTACCTACCATCTGCTGACAGCATGAAATTTGAGATT 999  
Qy 481 ggaagagctcttacttggtgcatattctctcctggttctccctgatacagtgaaatctc 540  
|||||  
Db 1000 GGAAGAGCTCTTACTTGGGCTATTTCTTCTCCCTGATAGCTGGAATCATC 1059  
Qy 541 ctctgcttctcctgctcatcccaagaagaatcgtctccaactactacagtgctcacaagcc 600  
|||||  
Db 1060 CTCTGCTTTTCTCTGCTCATCCAGAGAAATCGCTCCAACTACTAGATGCTTACCAAGCC 1119  
Qy 601 caactctctgcaacaagagctctcacaagcctggtgcaactcctccaaagtcagaagtgag 660  
|||||  
Db 1120 CAACCTCTTGGCACAAAGAGCTCTCCAAAGGCTGTGCTCAACTCTCCCAAAGTCAAGAGTGAG 1179

QY 661 ttcaattctacagcctgacagggatgtg 690  
|||||  
Db 1180 TTCAATTCTACAGCTGACAGGGTATGTG 1209  
|||||  
RESULT 6  
AXI36129  
LOCUS AXI36129 2863 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 51 from Patent EPI067182.  
ACCESSION AXI36129  
VERSION AXI36129.1 GI:14272537  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Plrimates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2863)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.  
TITLE Secretory protein or membrane protein  
JOURNAL Patent: Ep 1067182-A 51 10-JAN-2001;  
Helix Research Institute (JP)  
FEATURES  
source 1..2863  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
231..923  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39707.1"  
/db\_xref="GI:14272538"  
/translation="MASLGLQVGYILGLLGLTLVAMLLPSMKTSSYVGASIVTAV  
GFSKGLMECATSTGTQCDIYSLTGLPADIAQAAMVTSASISLACIIISVVG  
RCVFCQESRAKDRVAVGGVFLGLGLGFIPVANNLHGLIRDFYPLVPDSMKFEI  
GEALYGLITSLSFLIAGIILCFSCSSQNRNSNYDAYQAQPLATRSRPGQPPKYK  
SEFNSYSLGIV"  
BASE COUNT 670 a 794 c 711 g 688 t  
ORIGIN  
Query Match 100.0%; Score 690; DB 6; Length 2863;  
Best Local Similarity 100.0%; Pred. No. 1.5e-191;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctctcttggcctcccaactgtgggtacacatcctagcctcttcttggccttttgggc 60  
|||||  
Db 231 ATGGCCTCTTGGCCCTCAACTTGTGGGTACATCTAGGCCCTCTGGGCTTTTGGGC 290  
|||||  
QY 61 acaactgttgcattgtctctccagctggaacaaagttcttattgtcgtgcccagcatt 120  
|||||  
Db 291 ACATGTTGCCATGCTGCTCCCGAGCTGGAAACAAAGTTCTTATGTGGTCCAGCATT 350  
|||||  
QY 121 gtacacagattggcttctccaaggcctctggatggaattgtccacacacagcagc 180  
|||||  
Db 351 GTGACAGCAGTGGCTTCTCCAGGGCCTCTGGATGGAATGTGCACACAGCAGCAGGC 410  
|||||  
QY 181 atcaccagtgtagactctatagcacctcttggcctgcccctgacatccagcctgcc 240  
|||||  
Db 411 ATACCCAGTGTGACATCTATAGCACCTCTGCGGCTGCCCGGTGACATCCAGCTGCC 470  
|||||  
QY 241 caggccatgatgttgacatccagtgcaatctctccctggcctgattctctctgtgtg 300  
|||||  
Db 471 CAGGCCATGATGTTGACATCCAGTGCAATCTCCCTCCCTGGGCTGACATCTCTGTGTG 530  
|||||  
QY 301 ggcattagatgcacagctctcttccagggaatcccgagccaaagacagatggcggtagca 360  
|||||  
Db 531 GGCATGAGATGCACAGTCTTCTGCGAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 590  
|||||  
QY 361 ggtgagatctttttcatcttggaggcctcctggattcattctctgttgcctggaattct 420  
|||||  
Db 591 GGTGGAGTCTTTTTCATCTTGGAGGCTCTCTGGGATTCATCTCTGTGTGCTGGAATCTT 650  
|||||  
QY 421 catgggatactcagggacttctactaccactggtgctgacagcatgaaatttgagatt 480  
|||||

Db 651 CATGGATCCTACGGGACTTCTACTCACCACCTGGTGCCTGACAGCATGAATTTGAGATT 710  
|||||  
QY 481 ggagagcgtcttacttgggcattatttctccctgttctccctgatactgctggaatcattc 540  
|||||  
Db 711 GGAGAGGCTCTTACTTGGGCATATTCTTCTCCTGTTCTCCTGATAGTGAATCATC 770  
|||||  
QY 541 ctctgctttctctgctcattcccaagagaatacgtctcaactactacgctaccagcc 600  
|||||  
Db 771 CTCTGCTTTTCTGCTCATCCAGAGAAATGCTCCAACCTACTACTACGATGCTACCAAGCC 830  
|||||  
QY 601 caacctcttgccacaaggagctctccaaagcctggtcacaacctccccaaagtcaagagtga 660  
|||||  
Db 831 CAACCTCTTTCACCAAGAGAGCTCTCCAAAGGCTGGTCAACCTCCCAAGTCAAGAGTGAG 890  
|||||  
QY 661 ttcaattctacagcctgacagggatgtg 690  
|||||  
Db 891 TTCAATTCTACAGCTGACAGGGTATGTG 920  
|||||  
RESULT 7  
AL158821  
LOCUS AL158821 109465 bp DNA linear PRI 08-FEB-2002  
DEFINITION Human DNA sequence from clone RPI-75H8 on chromosome Xq22.3-23  
Contains the CLDN2 gene for claudin 2, the gene for a novel protein  
similar to KIAA0136, a eukaryotic translation elongation factor 1  
alpha 1 (EEF1A1) pseudogene and a CpG island, complete sequence.  
ACCESSION AL158821  
VERSION AL158821.16 GI:14329875  
KEYWORDS HTG; claudin; CLDN2; CpG island; EEF1A1; KIAA0136.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Plrimates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 109465)  
AUTHORS Heath, P.  
TITLE Direct Submission  
JOURNAL Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
On Jun 8, 2001 this sequence version replaced gi:14160908.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em.: EMBL; Sw.:  
SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/projects/c\_elegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human  
chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/ChrX  
RPI-75H8 is from the library RPI-1 constructed by the group of  
Pietter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone RPI-75H8  
It may be shorter because we sequence overlapping sections only  
once, except for a short overlap.  
The true left end of clone RPI-75H8 is at 1 in this sequence. The  
true left end of clone RPI3-383K5 is at 109366 in this sequence.  
FEATURES Location/Qualifiers



QEOTVEKKMPMENENHOVFNPPKILITQVEMAGLNKNTIGYEGIHSPSLPSCGEE  
SRSPSLQKPLDSSVLFSSKYKWLGEPEKRRRLQNEHTTSLDYSMPAPRRVE  
APVAYPEGENSHDKSSSERTPPYLFPEYPEASANTGONREVSILYIPKADQROGSLI

Query Match 100.0%; Score 690; DB 9; Length 109465;  
Best Local Similarity 100.0%; Pred. No. 1.9e-191;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctccaaacttggtggtacatctcttagcctctcttggtggtc 60  
|||||  
Db 5019 ATGGCCCTCTCTGGCCCTCCAACTTGTGGGCTACATCTTAGGCTTCTGGGCTTTGGGC 5078  
QY 61 acactgttccatgctgctcccccagctggaacaaagtcttattgtcgtgccaagatt 120  
|||||  
Db 5079 ACATGGTGTGCAATGCTGCTCCCCAGCTGGAAACAAAGTTCTATGTGGTGCCAGCAAT 5138  
QY 121 gtgacacagttggtctctccaaaggcctctggatggaatgtgccacacacacagcaggc 180  
|||||  
Db 5139 GTGACACAGTTGGCTTCTCCAAAGGCTCTGGATGGAATGTGCCACACACACAGCAGGC 5198  
QY 181 atcacccagttgacatctatatagaccccttctggcctgcccgtgacatccaggctgcc 240  
|||||  
Db 5199 ATCACCAGTGTGACATCTATAGCACCTTCTGGGCTGCCCCGTGACATCCAGGCTGCC 5258  
QY 241 caggccatgattgacatccagtgcaatctccctcctggtcgtgacattctctgtggtg 300  
|||||  
Db 5259 CAGGCCATGATGTTGATCCAGTGCATTCCTCCCTGCGCTGCATTTATCTGTGTGTG 5318  
QY 301 ggcattgagatgcagctctcttcgcagggaatcccgagccaaagacagagtgycggtagca 360  
|||||  
Db 5319 GGATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 5378  
QY 361 ggtggagttcttttcatctcttggtggtcctctgggtattcctctgttgcctggaattctt 420  
|||||  
Db 5379 GGTGGAGTCTTTTTCATCTCTTGAGGCTCTCTGGGATTCATTTCTGTGCTGGAAATCTT 5438  
QY 421 catgggactcaggggactctctactcaccactggtgctgacagcatgaaatttgagatt 480  
|||||  
Db 5439 CATGGGATCTTACGGGACTTCTACTCACCACCTGGTGCCTTGACAGCATGAAATTTGAGATT 5498  
QY 481 ggagaggtctttacttgggcatctattctccctgttctccctgtagctggaatcattc 540  
|||||  
Db 5499 GGAGAGGCTCTTTACTTGGGCAATTTCTTCCCTGTCTCCCTGATAGCTGGGAATCATC 5558  
QY 541 ctctgctttctgctctatccacagagaatacgtcccaactactcagatgctcacaagcc 600  
|||||  
Db 5559 CTCTGCTTTTCTGCTCATCCACAGAGAAATCGCTCCAACTACTACGATGCCCTACCAAGCC 5618  
QY 601 caacctcttgccaaagagctctccaaaggcctggttcaacctcccaagtcagaagttag 660  
|||||  
Db 5619 CAACCTCTTGGCCACAAGGAGCTCTCCAAGGCTGGTCAACCTCCCAAGTCAAGAGTGAG 5678  
QY 661 ttcaattctctacagctgcaggggtatgtg 690  
|||||  
Db 5679 TTCAATTCCTACAGCTGACAGGGTATGTG 5708

RESULT 8  
AX286822 1441 bp DNA linear PAT 21-NOV-2001  
LOCUS  
DEFINITION Sequence 1 from Patent WO0180879.  
ACCESSION AX286822  
VERSION AX286822.1 GI:17048854

KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Vinals Y de Bassols, C.G., Gaulis, S.R., Coche, T.G. and Orntoft, T.A.  
TITLE Colorectal cancer vaccines and diagnosis  
JOURNAL Patent: WO 0180879-A 1 01-NOV-2001;  
SMITHKLINE, BEECHAM BIOLOGICALS (S.A.)

FEATURES  
source Location/Qualifiers  
l. .1441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 337 a 399 c 357 g 345 t 3 others  
ORIGIN

Query Match 99.9%; Score 689.6; DB 6; Length 1441;  
Best Local Similarity 99.9%; Pred. No. 2e-191;  
Matches 689; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctccaaacttggtggtacatctcttagcctctcttggtggtc 60  
|||||  
Db 77 ATGGCCCTCTCTGGCCCTCCAACTTGTGGGCTACATCTTAGGCTTCTGGGCTTTGGGC 136  
QY 61 acactgttggcatgctgctcccccagctggaacaaagtcttattgtcgtgccaagatt 120  
|||||  
Db 137 ACATGGTGTGCAATGCTGCTCCCCAGCTGGAAACAAAGTTCTATGTGGTGCCAGCAAT 196  
QY 121 gtgacacagttggtctctccaaaggcctctggatggaatgtgccacacacacagcaggc 180  
|||||  
Db 197 GTGACACAGTGGCTTCTCCAAAGGCTCTGGATGGAATGTGCCACACACACAGCAGGC 256  
QY 181 atcacccagttgacatctatatagcacccttctggcctgcccgtgacatccaggctgcc 240  
|||||  
Db 257 ATCACCAGTGTGACATCTATAGCACCTTCTGGGCTGCCCCGTGACATCCAGGCTGCC 316  
QY 241 caggccatgattgacatccagtgcaatctccctcctggtcgtgacattctctgtggtg 300  
|||||  
Db 317 CAGGCCATGATGTTGATCCAGTGCATTCCTCCCTGCGCTGCATTTATCTGTGTGTG 376  
QY 301 ggcattgagatgcagctctcttcgcagggaatcccgagccaaagacagagtgycggtagca 360  
|||||  
Db 377 GGATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGCA 436  
QY 361 ggtggagttcttttcatctcttggtggtcctctgggtattcctctgttgcctggaattctt 420  
|||||  
Db 437 GGTGGAGTCTTTTTCATCTCTTGAGGCTCTCTGGGATTCATTTCTGTGCTGGAAATCTT 496  
QY 421 catgggactcaggggactctctactcaccactggtgctgacagcatgaaatttgagatt 480  
|||||  
Db 497 CATGGGATCTTACGGGACTTCTACTCACCACCTGGTGCCTTGACAGCATGAAATTTGAGATT 556  
QY 481 ggagaggtctttacttgggcatctattctccctgttctccctgtagctggaatcattc 540  
|||||  
Db 557 GGAGAGGCTCTTTACTTGGGCAATTTCTTCCCTGTCTCCCTGATAGCTGGGAATCATC 616  
QY 541 ctctgctttctgctctatccacagagaatacgtcccaactactcagatgctcacaagcc 600  
|||||  
Db 617 CTCTGCTTTTCTGCTCATSCCAGAGAAATCGCTCCAACTACTACGATGCCCTACCAAGCC 676  
QY 601 caacctcttgccaaagagctctccaaaggcctggttcaacctcccaagtcagaagttag 660  
|||||  
Db 677 CAACCTCTTGGCCACAAGGAGCTCTCCAAGGCTGGTCAACCTCCCAAGTCAAGAGTGAG 736  
QY 661 ttcaattctctacagctgcaggggtatgtg 690  
|||||  
Db 737 TTCAATTCCTACAGCTGACAGGGTATGTG 766

RESULT 9  
AF358907  
LOCUS AF358907 953 bp mRNA linear MAM 08-MAY-2001  
DEFINITION Canis familiaris integral membrane protein claudin-2 mRNA, complete cds.  
ACCESSION AF358907  
VERSION AF358907.1 GI:13991612  
KEYWORDS  
SOURCE dog.  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 953)  
AUTHORS Furuse, M., Furuse, K., Sasaki, H. and Tsukita, S.  
TITLE Conversion of Zonulae Occludentes from Tight to Leaky Strand Type by Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells  
JOURNAL J. Cell Biol. 153 (2), 263-272 (2001)  
MEDLINE 21206012  
PUBMED 11309408  
REFERENCE 2 (bases 1 to 953)  
AUTHORS Furuse, M., Furuse, K., Sasaki, H. and Tsukita, S.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-2001) Cell Biology, Kyoto University, Faculty of Medicine, Yoshida-Konoie, Sakyo, Kyoto 606-8501, Japan  
FEATURES  
source  
1.953  
/organism="Canis familiaris"  
/db\_xref="taxon:9615"  
/sex="male"  
/tissue\_type="liver"  
59..751  
/note="localizes to tight junctions"  
/codon\_start=1  
/product="Integral membrane protein claudin-2"  
/protein\_id="AAK51433.1"  
/db\_xref="GI:1391613"  
/translation="MASLGLQVGYILGLGLGLIVAMHLSWRTSSVYGISIVAV  
GFSKIMBECATHSTGIVOCIDLYSTILGLPADIQAAQAMWVSSAISLACTISVVM  
RCTVFCQDSRARHDLAVGVGFIIIGLGFIPVAMNHLIDFYSPLVPDSMKREI  
GEALYIGLISLFSIVAGIILICFSCPLQGNRSYDYSQAQPLATRGSPRPGPPAK  
SEFNSYSLTGYV"  
BASE COUNT 195 a 289 c 265 g 204 t  
ORIGIN

Query Match 81.4%; Score 562; DB 4; Length 953;  
Best Local Similarity 88.4%; Pred. No. 5.8e-154;  
Matches 610; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctcaactgtggtgctacaccccttgagccttgaggccttggc 60  
DB 59 ATGGCCTCTCTCGGCTCCCAACTGTGAGGCTACATCTCTAGGCTCTGCGGCTGTGGGC 118  
QY 61 acactgttgccctgctgctcccaactgtgaaacaaagtcttatgtcgggtgccaagc 120  
DB 119 ACCCTGTGGCCATGCTGCTTCCAGCTGGCGAACAAGCTCTACGTTGGTACAGCATC 178  
QY 121 gtgacagcagttgctcttcccaaggcctctgagtgaatgtgccaacacagcagc 180  
DB 179 GTGACGGCGGTGGCTTCTCCCAAGGGGCTCTGATGAGAGTGGCCACACAGCAGCAGC 238  
QY 181 atcaccagttgacatctatagcaaccctctggtgctgcgcgtgacatccagctgccc 240  
DB 239 ATTAACCCAGTGTGACATCTACAGCACACCCTCTAGGCTCTGCTGTGACATCCAGCTGCC 298  
QY 241 caggcctatagtgacatccatgacatctcctcctgctgcgcgtacatctctgtgtg 300  
DB 299 CAGGCCATGATGGTGAATCCAGTCCAGTCTTCTGTTGGCTGACATGCTCTGTGGTG 358  
QY 301 ggcattgagatgacagctctctcgcaggaatcccgagcaacagagtgagcgtagca 360  
DB 359 GGCATGAGATGCACTGCTCTTCTCCAGAGATCCCGAACCAGAACAGACATGGCGGTGGTG 418  
QY 361 ggtgagagcttttcaatcccttgaggcctctgagatcattctcgtgttgcttgaaatct 420  
DB 419 GGTGAGAGCTTTCATTCATTTGAGAGGCTCTGAGGCTTCATCCCGTTCGCTGGAACCTT 478  
QY 421 catgagatcttaaggagactctacacacagctgctgcaagcatgaaatttgagatt 480  
DB 479 CACGGGATCTCGGAGACTTCTACTCCGCCCTGTACCCCATAGCATGAGATTGAGATC 538  
QY 481 ggaagagcctcttaacttgaggcatattctcctgttctccctgataagctgaaatc 540  
DB 539 GGAAGAGCTCTTACTTGGGCAATTAATTCCTCTGTTCTCCCTGGTAGCTGGCATCATC 598

QY 541 ctctgctttctctgctatcccgagaaatgctctccactactagctaccagacc 600  
DB 599 CTCTGCTTTCTCTCCCACTCCAGGAAATGCTCCGACTACTATGACTCTACAGGCC 658  
QY 601 caactcttgccacaaggagctctcccaaggcctgtgtaacctcccaagatgaagtgag 660  
DB 659 CAGCCCTTGGCACTAGAGGCTCTCCAGGCCGGGTCAACGCCCAAGCCCAAGAGCAG 718  
QY 661 ttcaattctacagcctgacaggtatgtg 690  
DB 719 TTTAACTCTTACAGCTGACAGGATGTGTG 748  
RESULT 10  
AX286824  
LOCUS AX286824 779 bp DNA linear PAT 21-NOV-2001  
DEFINITION Sequence 3 from Patent WO0180879.  
ACCESSION AX286824  
VERSION AX286824.1 GI:17048855  
KEYWORDS  
SOURCE  
MUS SP.  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Vinals y de Bassols, C.G., Gaulis, S.R., Coche, T.G. and Ortoft, T.A.  
TITLE Colorectal cancer vaccines and diagnosis  
JOURNAL Patient: MO 0180879-A 3 01-NOV-2001;  
SMITHKLINE BEECHAM BIOLOGICALS (S.A.)  
FEATURES  
source  
1..779  
/organism="Mus sp."  
/db\_xref="taxon:10095"  
BASE COUNT 152 a 205 c 216 g 206 t  
ORIGIN

Query Match 79.1%; Score 546; DB 6; Length 779;  
Best Local Similarity 87.0%; Pred. No. 2.9e-149;  
Matches 600; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctcaactgtggtgctacaccccttgagccttgaggccttggc 60  
DB 48 ATGGCTCTCTCGGCTCCCAACTGTGAGGCTACATCTCTAGGCTCTGCGGCTGTGGGC 107  
QY 61 acactgttgccctgctgctcccaactgtgaaacaaagtcttatgtcgggtgccaagc 120  
DB 108 ACATTCATTTGCCATGCTCTTCCCAACTGGCGAACAAGTTCCTATGTGGTCCAGCATT 167  
QY 121 gtgacagcagttgctcttcccaaggcctctgagtgaatgtgccaacacagcagcagc 180  
DB 168 GTGACGGCGGTGGCTTTCACAGGCGCTCGAGATGAGTGTGGCACACACAGCAGCAGC 227  
QY 181 atcaccagttgacatctatagcaaccctctggtgctgcgcgtgacatccagctgccc 240  
DB 228 ATCAACCCAGTGTGACATCTACAGCACACCCTTATGAGACATCTCTGCTGTGAGCTGCC 287  
QY 241 caggcctatagtgacatccatgacatctcctcctgctgcgcgtacatctctgtgtg 300  
DB 288 CAGGCCATGATGGTGAATCCAGTCCAGTCTTCTGTTGGCTGACATGCTCTGTGGTG 347  
QY 301 ggcattgagatgacagctctctcgcaggaatcccgagcaacagagtgagcgtagca 360  
DB 348 GGCATGAGATGCACTGCTCTTCTCCAGAGATCCCGAACCAGAACAGACATGGCGGTGGTG 407  
QY 361 ggtgagagcttttcaatcccttgaggcctctgagatcattctcgtgttgcttgaaatct 420  
DB 408 GGTGAGAGCTTTCATTCATTTGAGAGGCTCTGAGGCTTCATCCCGTTCGCTGGAATCTT 467  
QY 421 catgagatcttaaggagactctacacacagctgctgcaagcatgaaatttgagatt 480  
DB 468 CATGGCAATCTTCGAGACTTCTACTTCGCGCTGTTCCTGTACAGCATGAAATTTGAGATT 527

```
QY 481 ggagaggctcttacttgggcatatttctccctgttctccctgatacagtggaatc 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 GGAGAGGCTCTGACTTGGGATCATCTACGCCCTGTTTCTTGGTAGCGAGTCATC 587
QY 541 ctctgcttttctctgctcatccagagaaatcgctccaaactactacatgcctaccaagcc 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 CTTTGGCTTTCTGCTCGCCCCAGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 647
QY 601 caaccttttcccaagaagagctctccaaagcctgggtccaaacctcccaaaagtccaagtgag 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGTCTCAACAGCCCAAGCAAGAGTGAG 707
QY 661 ttaaatccctacagctgacaggttatgtg 690
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 TTCAACTCATACAGCCCTGACTGGGTATGTG 737

RESULT 11
E31591          B31591          791 bp  DNA  linear  PAT 07-FEB-2001
LOCUS          Tight junction-constituting membrane protein claudin family.
DEFINITION
ACCESSION      E31591
VERSION        E31591.1 GI:13018519
KEYWORDS       JP 2000032984-A/2.
SOURCE         Mus sp.
ORGANISM       Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Svolchiro,T.
Tight junction-constituting membrane protein claudin family
Patent: JP 2000032984-A 2 02-FEB-2000;
EISAI CO LTD
OS             Mus sp. (mouse)
PN             JP 2000032984-A/2
PD             02-FEB-2000
PF             26-JUN-1998 JP 1998179847
PR
PI             SYOICHIRO TSUKITA
PC             C12N15/09,C07K14/47,C07K16/18,C12N5/10,C12P21/02,C12Q1/02,PC
C12Q1/68,
PC             G01N33/15,G01N33/53,G01N33/577//C12P21/08,(C12N5/10,C12R1:91),
PC             (C12P21/02,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC
Strandedness: Double;
CC             Topology: Linear;
FH             Key Location/Qualifiers
FT             CDS 48..740.

FEATURES
source         1..791
                /organism="Mus sp."
                /db_xref="taxon:10095"
BASE COUNT     155 a 207 c 220 g 209 t
ORIGIN

Query Match    79.1%; Score 546; DB 6; Length 791;
Best Local Similarity 87.0%; Pred. No. 2.9e-149;
Matches 600; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1 atggcctcttggcctccaaactgtgtgggtacatctcctagcctctgtggccttttgggc 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 ATGGCTCCCTTGGCTTCAACTGGTGGCTTACATCCTAGGCCCTTTTGGGCTGTAGGC 107
QY 61 acatgggttccatgctgtcccccagctggaaacaaagttcttatgtcgggtccagcatt 120
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 ACATCCATTTGGCTTCCAACTGCTTCCAACTGGCGAAGAGTTCCTATGTGGTCCAGCAT 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 gtacacagcagtgtgcttctccaaagggcctctggatggaatgtcacacacagcacagc 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 GTACGCGGCTTGGCTTTTCCAGGGCTCTGGATGAGTGTGGCGACACACAGCACAGC 227
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 atcaccagtgtagacatctatagcacccttctggcctgcccgtgacatccagcgtgcc 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 228 ATCACCCAGTGGCATATCTACAGTACCCCTTTTAGGACTTCTCTGCTGACATCCAGGCTGCC 287
QY 241 cagggccatgatggtgacatccagtgcaatctctccctggcctgcattatctctgtgggtg 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 CAGGCCATGATGTCAGCTCCAGTGCATGTCCTCGCTGGCTTGTATTTATCTCTCTGTGGTG 347
QY 301 ggcataagatgacagtccttctgccaaggaatcccgagagcaaaagacagagtgcggtagca 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 GGCATGAGATGCACCGCTGTCTGCCAGGATTCCTCGAGCTAAAGACAGAGTGGCTAGTG 407
QY 361 ggtgagtgcttttctcatctctgagagcctcctggggttcattcctgttgcctggaatctt 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 GTGGAGTCTTTTTCATCTCTGTGGCATCTCTCAGCCCTGTTTCTTGGTAGCGGAGTCATC 467
QY 421 catggatcctacgggacttctactcaccactggtgctcgacagcatgaaatttgagatt 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 CATGGCATCTCTCGGACCTTCTACTCGCCGCTGGTTCCTGACAGCATGAAATTTGAGATT 527
QY 481 ggagaggtcttttacttgggcatatttctccctgttctccctgatacagtggaatc 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 GGAGAGGCTCTGACTTGGGATCATCTCAGCCCTGTTTCTTGGTAGCGGAGTCATC 587
QY 541 ctctgcttttctgctcatccagagaaatcgctccaaactactacatgcctaccagcc 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 CTTTGGCTTTCTGCTCGCCCCAGGCAATCGTACCAACTACTATGATGGCTACCAGGCC 647
QY 601 caacctttgccaagaagagctctccaaagcctgggtccaaacctcccaagccaagtgag 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 CAGCCTCTTGGCACTAGGAGCTCTCCAAGATCTGCTCAACAGCCCAAGCAAGAGTGAG 707
QY 661 ttaaatccctacagctgacaggttatgtg 690
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 TTCAACTCATACAGCCCTGACTGGGTATGTG 737

RESULT 12
AF072128        791 bp  mRNA  linear  ROD 21-JUL-1998
LOCUS          Mus musculus claudin-2 mRNA, complete cds.
DEFINITION
ACCESSION      AF072128
VERSION        AF072128.1 GI:3335183
KEYWORDS       house mouse.
SOURCE         Mus musculus.
ORGANISM       Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 791)
AUTHORS        Furuse,M., Fujita,K., Hiiiragi,T., Fujimoto,K. and Tsukita,S.
TITLE          Claudin-1 and -2: novel integral membrane proteins localizing at
                tight junctions with No sequence similarity to occludin
JOURNAL        J. Cell Biol. 141 (7), 1539-1550 (1998)
MEDLINE        98311639
REFERENCE      2 (bases 1 to 791)
AUTHORS        Furuse,M., Fujita,K., Hiiiragi,T., Fujimoto,K. and Tsukita,S.H.
TITLE          Direct Submission
SUBMITTED      (15-JUN-1998) Cell Biology, Kyoto University Faculty of
                Medicine, Sakyo-Ku, Kyoto 606-8501, Japan
FEATURES       Location/Qualifiers
source         1..791
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                48..740
                /note="integral membrane protein localizing at tight
                junctions"
                /codon_start=1
                /product="claudin-2"
                /protein_id="AAC27079.1"
                /db_xref="GI:3335184"
                /translation="MASLGVQLVGLILGLLGLTSLIAMLLPNWRTSSVGNISVTAV
                GFSKGLMWEKATHSTGTCDDIYSLDGLPADIAQAAMWVTSSAMSLACILISVVG
                RCTVFCQDSKARVRVAVGGVFFILGGILGFIPLVANNHLGFLVLPDMSKFEI
                GEALYGLISALFSLVAGVILCFSCSPQGNRTNYDGYQAOPLATRSSPSRAQPKAK
                SEFNSYSLTGYV"
```

BASE COUNT	155 a	207 c	220 g	209 t
ORIGIN				

Query Match	79.1%	Score 546;	DB 10;	Length 791;
Best Local Similarity	87.0%	Pred. No. 2.9e-149;		
Matches 600;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0;

OY	1	atgagctctcttggagctccaacttlytgggctaacatcccttaaggctctcttgggcttttgggc	60
Db	48	ATGGCCTCCCTTGGGGTTCACATGGTGGGCTTACATCTTAGGCTTTTGGGCTGTTAAGC	107
OY	61	aactatgttgcacatgcctctcccaagcttggaaaacaagttctatctgtcgttgcagcatt	120
Db	108	ACATCATTTGGCATCTGCTGCTCCCAACTGGGCAACGAGAGTTCTATGTGGTGCAGCATT	167
OY	121	gttaagaagaatttggcttcccaaggctcttggatgtggaattgtgcacacaaagcacaagc	180
Db	168	GTGACGGGGGTTGGCTTTTCCAAAGGGCCCTTGGAATGGATGTGGCAACACAGCACAGGC	227
OY	181	atcacccagctgtgaacatctatagaacacctctcgggaccttgcgcctgtgaacatcaagctgcc	240
Db	228	ATCACCCAGTGGCATATCTACAGTACCCCTTTTAGGACTTCGTGCTGACATCCAGGCTGCG	287
OY	241	caaggcacaatgtgtgaacatcaatgcagatctctcccttgccttgcattatctctgtgtgtg	300
Db	288	CAGGCGCATGATGGTACGCGCATGCTCAATGTCTCTGCTGGCTTGATATATCTCTGTGGTG	347
OY	301	ggcattgagatgcacagctctctcgcaggaaatccgagcccaagaacagaagtgtgcgtlaaga	360
Db	348	GGCATGAGATGACCGGTGTCTGCCAGATTTCTCAGACTAAGGACACAGATGGCTGTAGTG	407
OY	361	ggttgagctcttttcatcccttggaggacctccttggatctcaatctctgttgccttgaactct	420
Db	408	GGTGAAGCTTTTTCATCTTGCTGGCATCTCGGGCTTATCCAGATTGCTGTGAATCTT	467
OY	421	catggatctccacggagactcttaccacacatcgtgcctcggagaagatgaatattgaatt	480
Db	468	CATGGCATCTCTCGGGAATCTTACTTCGCGCGCTGGTCTCTGACACGATGAATATTAGATT	527
OY	481	ggagagagctcttacttgggaattattcttccctgttctccctgatagacttgaatcatc	540
Db	528	GGAGAGGCTCTGTACTTGGGCATCTACATCTACAGCCCTGTGTTCTTGTGGTCCGGAGTCATC	587
OY	541	ctctgcctttctctgtcatcccaaggaaatcgtctcaactactaagaatgacctacaagc	600
Db	588	CTTTCCTTTCTGTCTCGCCCCAGGGGCATCTACTCAACTACTATGATGTGGTCAACAGGCC	647
OY	601	caacctcttgcacaagaagactctctcaagagccttgttcaacctccccaagaatcaagagttag	660
Db	648	CAGCTCTTGCCACTAGGAGCTCTCCAGATGTGCTCAACACAGCCCAAGGACCAAGAGTGAG	707
OY	661	ttcaattcctacaagcctgaacaggtatggtg 690	
Db	708	TTCAACTCATACAGCGCTGACTGGGTATATGTG 737	

RESULT	13				
LOCUS	BC015252				
DEFINITION	BC015252	2628 bp	mRNA	linear	ROD 29-OCT-2001
ACCESSION	BC015252				
VERSION	BC015252.1				
KEYWORDS	MG.				
SOURCE	house mouse.				
ORGANISM	Mus musculus.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2828)				
AUTHORS	Strausberg,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-OCT-2001)				National Institutes of Health, Mammalian

REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.inl.gov>  
Series: IRAC Plate: 25 Row: h Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7710003.

FEATURES	Location/Qualifiers
source	1. .2828

```

/organism="Mus musculus"
/db_xref="locusid:12738"
/db_xref="taxon:10090"
/clone="MGC:18348 IMAGE:4223446"
/tissue_type="kidney, normal. 5 month old male mouse."
/clone_lib="NCI CGAP_K1d14"
/lab_host="DH10B"
/vec="Vector: PCMV-SPORT6"
126. .818

```

BASE COUNT	736 a	697 c	667 g	728 t
ORIGIN				

Query Match	79.1%;	Score 546;	DB 10;	length 2828;
Best Local Similarity	87.0%;	Pred. No. 3.1e-149;		
Matches 600;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0

Oy	1	atggctctctttggcttccaactgtggtggcacaatctcaagcctcttgggcttttggc	60
Db	126	atggctctcccttggggcttcaactggtgggctacatccttaggccttctttagc	185
Oy	61	aaactggttcacatgctgctcccccagctcggaaacaacttctatgtggtggcaactt	120
Db	186	acatccatttgcctatctctctgtcccaacttggcgaacgaattcttatgttgccacatt	245
Oy	121	gtacaagcagttgtgcttctccaagggtcctctgatatgtaatgtgcacacacagcagac	180
Db	246	gtgacggcggttgctgttttccaaaggccctctggatggagtgctgcacacacagcagac	305
Oy	181	atcacccagttgtgacatctatagcaacctctggtgcctgcgcgtgacatccaagctgcc	240
Db	306	atcacccagttgtgacatctatagcaacctctggtgcctgcgcgtgacatccaagctgcc	365
Oy	241	caaggcctatgattgatactccaagtgcaactctctccctggcctgcatctatctctgtgtg	300
Db	366	caaggcctatgattgatactccaagtgcaactctctccctggcctgcatctatctctgtgtg	425
Oy	301	ggcatggaatgacaagcttctcttccagaagaatccaggacaaagaagatggcggtagca	360
Db	426	ggcatggaatgacaagcttctcttccagaagaatccaggacaaagaagatggcggtagca	485

```
Qy 361 ggtgagctcttttcatctcttgaggcctcctgggattcattcctctgttgcctggaaattc 420
Db 486 GGTGAGCTCTTTTCATCTCTGTGGCATCCCGGCTTTATCCAGTTCTGTGGAAATCTT 545
Qy 421 catggatcctacgggaacttactcaccaactggctgcctgacagatgaaattgagatt 480
Db 546 CATGCATCTTCGGGACTTCTACTCGCGCTGGTTCCTGCAGCATGAAATTTGAGATT 605
Qy 481 ggaagagctcttactcttgaggattatttcttccctgttctccctgatagctggaatcatt 540
Db 606 GGAGAGGCTCTGTACTTGGCATCATCAGCCCTGTTTCTTGTAGCCGGAGTCAATC 665
Qy 541 ctctgcttttctctatcccaagaatacgtctcaactactacgactcctaccaagcc 600
Db 666 CTTTGCTTTTCTGCTCGCCCGGCAATCTGACCACTACTATGATGGCTACCAAGGCC 725
Qy 601 caactcttgccacagagctctccaaggcctggtcgaacctcccaaaagtcgaagtggag 660
Db 726 CAGCCTCTTGCCACTAGGAGCTCTCCAAGATCTGCTCAACACGCCCAACCAAGAGTGAG 785
Qy 661 ttcaattctacagctgacaggggtatgtg 690
Db 786 TTCAACTCATACAGCCTGACTGGGTATGTG 815

RESULT 14
AX136475
LOCUS AX136475 615 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 397 from Patent EP1067182.
ACCESSION AX136475
VERSION AX136475.1 GI:14272879
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 397 10-JAN-2001;
Helix Research Institute (JP)
FEATURES
source
1. .615
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 133 a 174 c 157 g 143 t 8 others
ORIGIN
Query Match 52.4%; Score 361.6; DB 6; Length 615;
Best Local Similarity 97.4%; Pred. No. 4.2e-95;
Matches 375; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 1 atggcctctcttgccctccaactgtgggtacatacctcctagcctctctgggcttttgggc 60
Db 231 ATGGCCTCTCTTGGCTCCAACTTGTGGCTACATCTAGGCTTCTGGGCTTTTGGGC 290
Qy 61 acactggttgcattgctctccacagctggaaacaagtcttctatgtcgggtccagcatt 120
Db 291 ACATGGTGTGCATGCTGCTCCCCAGCTGGAAACAAGTCTTATGTGGTGCCAGCAAT 350
Qy 121 gtgacagcagttggcttctccaaggcctctggtggaatgtgccacacacagcagc 180
Db 351 GTGACAGAGTTGGCTTCTCCAGGGCCTCTGGATGGAAATGTGCCACACACAGCAGGC 410
Qy 181 atcaccagtgtagacatataagcacccttctggcctgcccgctgacatccaggctgcc 240
Db 411 ATCACCAGTGTGACATCATATAACACCCTTCTGGCCCTGCCCGCTGACATCCAGGCTGCC 470
Qy 241 caggccatgattgtagacatccagtgcaatctcctcctggcctggaattctctgtggtg 300
Db 471 CAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCAATTATCTCTGTGGTG 530
```

```
Qy 301 ggcatagatgacagctcttctgccagg-aatcccagagccaaagacagagtgccggtagc 359
Db 531 GGCATGANATGCCACAGTCTTCTCCAGGNAATCCCGAACCAANACANAATGGCGGTANC 590
Qy 360 aggtggagctcttttcatcctctgga 384
Db 591 AGGTGGAAATCTTTTTCATCCTTGA 615

RESULT 15
HSAL32445
LOCUS Homo sapiens CLDN14 gene.
DEFINITION AJ132445
ACCESSION AJ132445
VERSION AJ132445.1 GI:4186039
KEYWORDS claudin-14; CLDN14 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Keen, T. J.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Keen T.J., Molecular Medicine Unit,
University of Leeds, Clinical Sciences Building, St James's
Hospital, Leeds, LS9 7TF, UK
REFERENCE
AUTHORS Keen, T. J. and Inglehearn, C. F.
JOURNAL Unpublished
COMMENT Related sequence AC000005.
FEATURES
Location/Qualifiers
source
1. .720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
1. .720
/gene="CLDN14"
1. .720
/gene="CLDN14"
/codon_start=1
/product="claudin-14"
/protein_id="CAAL0669.1"
/db_xref="GI:4186040"
/translation="MASTAVOLLGFLSLFGLMVGTLTITLPHWRRTAHVGNILTAV
SVLKLWMECVHSTGIYOCIVRSILALPODLQARALMWISLLSGIACACAVIGM
KCTRCAGKTPAKTTFALLGGLTFLLAGLCLWAVSWTNDVYQNFYVNFYPLPSGMKFEI
GOALYIGFISLSLSLIGLGLLCLSCODEAPYRYPYQAPPRATTTTANTAPAYQPPAAYK
DNRAPSVTSATHSGYRLNDYV"
BASE COUNT 126 a 270 c 193 g 131 t
ORIGIN
Query Match 31.3%; Score 215.8; DB 9; Length 720;
Best Local Similarity 61.8%; Pred. No. 2.8e-52;
Matches 343; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

Qy 1 atggcctctcttgccctccaactgtgggtacatacctcctagcctctctgggcttttgggc 60
Db 1 ATGGCCAGACGCGCCCTGCGACTTCTGGCTTCTGCTCAGCTTCTGGCATGTGGTGGC 60
Qy 61 acactggttgcattgctctccacagctggaaacaagtcttctatgtcgggtccagcatt 120
Db 61 AGTTGATCACCACCATCTCTGCCGACACTGCGGAGGACACGCCGACCCACACATC 120
Qy 121 gtgacagcagttggcttctccaaggcctctggtggaatgtgccacacacagcagc 180
Db 121 CTCACGCGCTGTCTCTACCTACCTGAAAGGCTCTGGATGGAGTGTGTGTGGCACACACAGGC 180
Qy 181 atcaccagtgtagacatataagcacccttctggcctgcccgctgacatccaggctgcc 240
Db 181 ATCTACCACTGGCAGATCTACCGATCCCTGCTGGCGCTGCCCAAGACCTCCAGGCTGCC 240
```

```

QY 241 caagccatgatgtgacatcccaagtgaatctccctccctgacctgacattatctctgtgtg 300
Db 241 CGGCCCCCTCATGGTCATCTCTGCTGCTCGGGCATAGCCTGCGCCGTCATC 300
QY 301 ggcattgagatgcacagctctctgccaagaaatcccgagccaaagaagagtgcggtagca 360
Db 301 GGGATGAAATGACAGCCCTGCGCCAAAGGCAACCCGCCAAGACACCTTTGCCATCCCTC 360
QY 361 gctggaagctctttcactccttgagagcctcctggaattcctctgtgacctggaatctt 420
Db 361 GCGGCGACCCCTTCATCTGCGCGGCTCTGTCATGGTGGCGCTCTCTGACACACC 420
QY 421 catggaatccttaaggaacttactcaacaactggtgccctgacagacatgaattgagatt 480
Db 421 AACGACCTGGTGACAGACTTCTACAAACCCGCTGCTGCCAGCGGCATGAACTTGAGATT 480
QY 481 ggaagagctcttacttggaattatttctccctgttccctgatatagctggaatcattc 540
Db 481 GCGCAGGCTCTGACCTGGGCTTCATCTCTGCTCTGCTCTGCTCATTTGGTGGACCCCTG 540
QY 541 ctctgcttttctctgc 555
Db 541 CTTTGGCTGTCTGTC 555

```

Search completed: June 2, 2002, 18:58:05  
 Job time: 11105 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:56:30 ; Search time 3664.46 Seconds  
(without alignments)  
88.397 Million cell updates/sec

Title: US-09-663-600A-2\_COPY\_1\_24  
Perfect score: 24  
Sequence: 1 gcauccuacuccaauucca 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	19.2	80.0	438	12	AQ887612 HS_5557_A
2	18.8	78.3	379	9	AV558212 AV558212
3	18.8	78.3	434	12	BH172929 SALK_0063
4	18.8	78.3	661	12	AG076330 Pan trogl
5	18.4	76.7	286	9	AV329061 AV329061
c 6	18.4	76.7	361	9	AI866251 w127f06.x
7	18.2	75.8	202	12	AQ248820 t16N9-Sp6
8	18.2	75.8	284	10	BM089677 503497 MA
c 9	18.2	75.8	321	9	AA153510 mq57b05.r
10	18.2	75.8	327	12	AZ442220 1M0234E19
11	18.2	75.8	343	9	AW789229 C00996-R
12	18.2	75.8	411	10	BF601807
c 13	18.2	75.8	431	9	AI645377 mc57b05.Y
14	18.2	75.8	462	9	BB860142 BB860142
c 15	18.2	75.8	511	12	AQ978003 RPCI-23-3
16	18.2	75.8	547	9	AW669920 113680 MA
17	18.2	75.8	577	12	BH033134 RPCI-24-3

18	18.2	75.8	605	10	BG689078
c 19	18.2	75.8	637	9	BB635865
c 20	18.2	75.8	666	9	BB208585
c 21	18.2	75.8	703	9	AI048533
22	18.2	75.8	816	12	AG146561
c 23	17.8	74.2	297	10	BE352981
c 24	17.8	74.2	428	10	WA3882
c 25	17.8	74.2	990	10	BG854007
c 26	17.6	73.3	250	9	AV021672
c 27	17.6	73.3	260	9	AW428024
c 28	17.6	73.3	265	12	A2920824
c 29	17.6	73.3	294	10	C20172
30	17.6	73.3	360	12	CNS01ZDQ
c 31	17.6	73.3	372	10	BI055852
32	17.6	73.3	401	10	BF991907
c 33	17.6	73.3	419	12	BH031217
c 34	17.6	73.3	429	9	AW484974
35	17.6	73.3	502	10	BM482975
c 36	17.6	73.3	507	10	BE405827
37	17.6	73.3	520	9	BE020754
c 38	17.6	73.3	567	10	BF231030
c 39	17.6	73.3	575	10	BJ073394
40	17.6	73.3	629	12	AO782163
c 41	17.6	73.3	643	10	BJ074708
c 42	17.6	73.3	666	10	BJ073088
43	17.6	73.3	673	12	AZ314706
c 44	17.6	73.3	679	10	BJ064086
45	17.6	73.3	685	10	BE254962

## ALIGNMENTS

### RESULT 1

#### AQ887612/c

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### COMMENT

AQ887612 438 bp DNA linear GSS 10-NOV-1999  
HS\_5557\_A2\_G10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=9325 Col=20 Row=M, DNA sequence.  
AQ887612  
AQ887612.1 GI:6343802  
GSS.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 438)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 9325 row: M column: 20  
Seq primer: Sp6  
Class: BAC ends  
High quality sequence stop: 438.  
Location/Qualifiers  
1. 438  
/organism="Homo sapiens"

```

/db_xref="taxon:3606"
/clone="Plate=9325 Col=20 Row=N"
/clone_lib="RPCr-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      154 a       34 c       67 g       162 t       1 others
ORIGIN

Query Match          80.0%; Score 19.2; DB 12; Length 438;
Best Local Similarity 62.5%; Pred. No. 7e+02;
Matches 15; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Oy      1 gaucacucccauccaaaucca 24
||| :||| :|||:||||:| |
Db      84 GCATCTACTTACCATTCCATTCTA 61

RESULT 2
AV558212              379 bp    mRNA     linear   EST 07-SEP-2000
LOCUS               AV558212 Arabidopsis thaliana green siliques Columbia Arabidopsis
DEFINITION           Thaliana cDNA clone S0092d03f 3', mRNA sequence.
ACCESSION             AV558212
VERSION               AV558212.1 GI:8729627
KEYWORDS              EST.
SOURCE                thale cress.
ORGANISM              Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE             1 Asamizu,E., Nakamura,Y., Sato,S. and Tsbata,S.
TITLE                A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
site-selected cDNA libraries
JOURNAL              DNA Res. 7, 175-180 (2000)
MEDLINE              20363093
COMMENT              Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
Source
location/qualifiers
1..379
organism="Arabidopsis thaliana"
strain="Columbia"
db_xref="taxon:3702"
clone="S0092d03f"
clone_lib="Arabidopsis thaliana green siliques Columbia"
tissue_type="green siliques"
note="Vector: pInscriptII SK-, Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      104 a       95 c       62 g       118 t
ORIGIN

Query Match          78.3%; Score 18.8; DB 9; Length 379;
Best Local Similarity 68.2%; Pred. No. 9.8e+02;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy      2 cauccaacucccauccaaaucc 23
||::||: ||::|||:||||:| |
Db      -245 CATCTCTCTCCCATCCAAAGTCC 266

RESULT 3
BH172929            434 bp    DNA     linear   GSS 03-OCT-2001
LOCUS              BH172929
```

DEFINITION	SALK_006364 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_006364, DNA sequence.
ACCESSION	BH172929
VERSION	BH172929.1 GI:15911190
KEYWORDS	GSS.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE	Alonso,J.M., Leisbe,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Heller,C., Kim,C.U., Jeske,A., Koesema,E., Meyers,M.C., Parker ,H., Predils,L., Shinn,P., Stevenson,D.R., Zimmerman,J. and Ecker ,J.R. A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
TITLE	Unpublished (2001)
JOURNAL	Contact: Joseph R. Ecker
COMMENT	Salk Institute Genomic Analysis Laboratory (SIGAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: eckeres@sal.k.edu This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 3' end of At1g01920. Class: TDNA tagged.
FEATURES	Location/Qualifiers  1..434 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_006364" /clone_lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna-protocols.html"
BASE COUNT	135 a               86 c               66 g               147 t
ORIGIN	
Query Match	78.3%; Score 18.8; DB 12; Length 434;
Best Local Similarity	68.2%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches	15; Conservative 5; Mismatches 2;
Oy	2 cauccacucaccaucaaucc 23
Db	:: :: :: :: :: ::  346 CATCCCTTCGCCATCAACTCC 367
RESULT	4
LOCUS	AG076330                        661 bp      DNA          linear     GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-070101.R, genomic survey sequence.
ACCESSION	AG076330
VERSION	AG076330.1 GI:16628132
KEYWORDS	GSS; GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM	BAC library clone:PTB-070101.R. Pan troglodytes
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan. Fujiyama,Y., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Torokii,Y., Watanabe,H. and Sakaki,Y. BAC end sequences of library PTB Unpublished 2 (bases 1 to 661)



JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbrrp/image/image.html  
Insert Length: 2439 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 360.

## FEATURES

source

1. .361  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2426147"  
/clone\_lib="NCI\_CGAP\_Utl"  
/tissue\_type="well-differentiated endometrial  
adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo df.  
Average insert size 1.75 kb. Life Technologies catalog #:  
11538-014"

BASE COUNT 102 a 65 c 90 g 104 t

## ORIGIN

Query Match 76.7%; Score 18.4; DB 9; Length 361;  
Best Local Similarity 70.0%; Pred. No. 1.4e+03;  
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 5 ccuacucccaucaaaucca 24  
||:||||:||||:||||:  
Db 216 CCTACTCCCATTCATTCCTCA 197

## RESULT 7

AQ248820 202 bp DNA linear GSS 06-OCT-1998  
LOCUS T16N9-Sp6 TAMU Arabidopsis thaliana genomic clone T16N9, DNA  
DEFINITION sequence.  
ACCESSION AQ248820  
VERSION AQ248820.1 GI:3698903  
KEYWORDS GSS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 202)  
Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and  
Ecker,J.

TITLE BAC End Sequences at ATGC

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
19104  
Tel: 215-898-9384  
Fax: 215-898-8780  
Email: jeckeratgenome.bio.upenn.edu  
Seq primer: Sp6  
Class: BAC ends.

## FEATURES

source

1. .202  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"

/clone="T16N9"  
/clone\_lib="TAMU"  
/sex="hermaphrodite"  
/note="Vector: BelOBACII; Site\_1: HindIII; Site\_2: HindIII  
; Produced by Rod Wing"

BASE COUNT 64 a 51 c 37 g 47 t 3 others  
ORIGIN

Query Match 75.8%; Score 18.2; DB 12; Length 202;  
Best Local Similarity 65.2%; Pred. No. 1.5e+03;  
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcauccuacucccaucaaaucc 23

||||:||||:||||:||||:  
Db 40 GCACCTACTTCTATCAATTCC 62

## RESULT 8

BM089677 284 bp mRNA linear EST 19-NOV-2001  
LOCUS 503497 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
DEFINITION  
ACCESSION BM089677  
VERSION BM089677.1 GI:17000305  
KEYWORDS EST.  
SOURCE cow.

ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.

## REFERENCE

## AUTHORS

1 (bases 1 to 284)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)

## JOURNAL

## MEDLINE

## COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 5 row: K column: 6

Seq primer: ATTTAGGTGACATATAG.

## FEATURES

source

1. .284  
Location/Qualifiers

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

BASE COUNT 46 a 104 c 94 g 40 t

## ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 284;  
Best Local Similarity 69.6%; Pred. No. 1.6e+03;  
Matches 16; Conservative 4; Mismatches 3; Indels 0; Gaps 0;



```

AW789229
LOCUS          343 bp      mRNA      linear      EST 01-MAY-2001
DEFINITION    C00996-R Lambda Zap, Stratagene Blumeria graminis f. sp. hordei
ACCESSION     CDNA clone C00996, mRNA sequence.
VERSION       AW789229.1 GI:13900826
KEYWORDS      EST.
SOURCE        Blumeria graminis f. sp. hordei.
ORGANISM      Blumeria graminis f. sp. hordei.
REFERENCE     1 (bases 1 to 343)
AUTHORS       Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and Oliver
              ,R.P.
TITLE         Gene identification in the fungal pathogen Blumeria graminis by
              expressed sequence tag analysis
JOURNAL       Unpublished (2000)
COMMENT       Contact: Rasmussen,S.W.
              Department of Yeast Genetics
              Carlsberg Laboratory
              10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark
              Tel: 45 3327 5230
              Fax: 45 3327 4766
              Email: sw@erc.dk
              High quality sequence stop: 343
POLYA-No.     Location/Qualifiers
              1..343
              /organism="Blumeria graminis f. sp. hordei"
              /db_xref="taxon:62688"
              /clone="C00996"
              /clone_lib="Lambda Zap, Stratagene"
              /cell_type="conidia"
              /lab_host="Hordeum vulgare"
              60 a 99 c 85 g 99 t

BASE COUNT    60 a 99 c 85 g 99 t
ORIGIN

Query Match   75.8%; Score 18.2; DB 9; Length 343;
Best Local Similarity 65.2%; Pred. No. 1.7e+03;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 cauccuacuccaaucaaucca 24
||||| :||||:||||| :|||
Db 29 CATCCCTATCCCATCCCAATGCCA 51

RESULT 12
LOCUS          411 bp      mRNA      linear      EST 25-APR-2001
DEFINITION    BF601807 MARC 380V Bos taurus cDNA 5', mRNA sequence.
ACCESSION     BF601807
VERSION       BF601807.1 GI:11699030
KEYWORDS      EST.
SOURCE        Cow.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovidae; Bovinae; Bos.
REFERENCE     1 (bases 1 to 411)
AUTHORS       Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
              Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
              ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
              Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
              Keele,J.W.
TITLE         Sequence evaluation of four pooled-tissue normalized bovine cDNA
              libraries and construction of a gene index for cattle
JOURNAL       Genome Res. 11 (4), 626-630 (2001)
MEDLINE       21180013
COMMENT       Contact: Smith TPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366

```

```

Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 42 row: K column: 6
Seq primer: ATTTAGTGACACTATAG.
              Location/Qualifiers
              1..411
              /organism="Bos taurus"
              /db_xref="taxon:9913"
              /clone_lib="MARC 380V"
              /tissue_type="pooled"
              /lab_host="DH10B"
              /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
              Library made from pooled tissue from marrow, alveolar
              macrophage, ovary, fetal semitendinosus muscle, and fetal
              longissimus muscle."
              64 a 131 c 133 g 83 t

BASE COUNT    64 a 131 c 133 g 83 t
ORIGIN

Query Match   75.8%; Score 18.2; DB 10; Length 411;
Best Local Similarity 69.6%; Pred. No. 1.7e+03;
Matches 16; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcauccuacuccaaucaauucc 23
||||| :||||:||||| :|||
Db 203 GCATCCAGCTCCCATCCCAACTCC 225

RESULT 13
LOCUS          431 bp      mRNA      linear      EST 29-APR-1999
DEFINITION    mq57b05.y1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:582801
              5', mRNA sequence.
ACCESSION     A1645377
VERSION       A1645377.1 GI:4723852
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE     1 (bases 1 to 431)
AUTHORS       NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
              National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
JOURNAL       Contact: Robert Strausberg, Ph.D.
              Email: cgaps-remail.nih.gov
COMMENT       This clone is available royalty-free through LLNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              This read is a RESEQUENCE of a previously sequenced mouse clone.
              This read has been verified (found to hit its original self in the
              correct orientation)
              Seq primer: -40RP from Gibco
              High quality sequence stop: 418.
              Location/Qualifiers
              1..431
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="IMAGE:582801"
              /clone_lib="Soares_thymus_2NbMT"
              /sex="male"
              /tissue_type="Thymus"
              /dev_stage="4 weeks"
              /lab_host="DH10B"
              /note="vector: pT73b-Pac (Pharmacia) with a modified
              polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

```



```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-339H5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
```

BASE COUNT 147 a 114 c 125 g 125 t

ORIGIN

```
Query Match 75.8%; Score 18.2; DB 12; Length 511;
Best Local Similarity 65.2%; Pred. No. 1.8e+03;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cauccuacucccauucca 24
|||:|:|:|:|:|:|:|
Db 275 CACCCTAATCCCTTCCAATTCCA 253
```

Search completed: June 2, 2002, 19:59:57  
Job time: 14607 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:54:15 ; Search time 108.28 Seconds  
(Without alignments)  
54.444 Million cell updates/sec

Title: US-09-663-600a-2\_COPY\_1\_24  
Perfect score: 24  
Sequence: 1 gcauccuacuccaaucca 24

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	46	4	US-08-905-223-2
2	24	100.0	46	4	US-09-247-155-2
3	24	100.0	47	4	US-08-905-223-1
4	24	100.0	47	4	US-09-247-155-1
5	24	100.0	65	3	US-08-930-102A-1
6	23	95.8	46	3	US-08-930-102A-5
7	23	95.8	46	3	US-08-930-102A-5
8	18.2	75.8	1312	2	US-08-841-349-6
9	18.2	75.8	6960	2	US-08-841-349-3
10	18.2	75.8	8176	2	US-08-841-349-5
11	16.2	67.5	250	3	US-08-301-162-13
12	16.2	67.5	250	4	US-09-461-240-13
13	16	66.7	305	1	US-08-170-294-5
14	16	66.7	305	2	US-08-664-855-5
15	16	66.7	305	3	US-09-049-289-5
16	16	66.7	2358	1	US-08-339-152A-15
17	16	66.7	2358	2	US-08-007-999B-2
18	16	66.7	2358	2	US-08-689-276A-2
19	16	66.7	2493	3	US-08-945-056-3
20	16	66.7	3827	1	US-08-170-294-6
21	16	66.7	3827	2	US-08-664-855-6
22	16	66.7	3827	2	US-08-718-751-1
23	16	66.7	3827	3	US-09-049-289-6
24	16	66.7	6671	1	US-08-280-443-1
25	16	66.7	6671	1	US-08-457-459-1
26	16	66.7	6671	1	US-08-555-678-1
27	16	66.7	6671	5	PCT-US95-02275-1

C 28	15.8	65.8	1664	1	US-07-863-169A-6	Sequence 6, Appli
C 29	15.8	65.8	1664	2	US-08-429-964-6	Sequence 6, Appli
C 30	15.8	65.8	1664	3	US-07-935-087-6	Sequence 6, Appli
C 31	15.8	65.8	1664	4	PCT-US93-08062-6	Sequence 427, App
C 32	15.6	65.0	567	4	US-09-385-982-427	Sequence 136, App
C 33	15.6	65.0	568	4	US-09-247-155-136	Sequence 15, Appli
C 34	15.6	65.0	569	3	US-09-109-204-15	Sequence 1, Appli
C 35	15.6	65.0	1843	3	US-08-718-738-1	Sequence 1, Appli
C 36	15.6	65.0	1843	4	US-09-221-844-1	Sequence 1, Appli
C 37	15.6	65.0	1843	5	PCT-US95-03323A-1	Sequence 1, Appli
C 38	15.6	65.0	2294	3	US-08-964-700A-1	Sequence 75, Appli
C 39	15.6	65.0	2411	3	US-09-188-930-75	Sequence 256, App
C 40	15.6	65.0	2411	3	US-09-188-930-256	Sequence 1, Appli
C 41	15.2	63.3	1288	4	US-09-296-754-1	Sequence 15, Appli
C 42	15	62.5	50	4	US-09-315-886C-15	Sequence 43, Appli
C 43	15	62.5	87	3	US-08-816-346-43	Sequence 43, Appli
C 44	15	62.5	87	3	US-09-335-411-43	Sequence 39, Appli
C 45	15	62.5	153	3	US-08-816-346-39	

## ALIGNMENTS

RESULT 1  
US-08-905-223-2  
; Sequence 2, Application US/08905223  
; Patent No. 6222029  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Jean-Baptiste D.  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Lacroix, Bruno  
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
; NUMBER OF SEQUENCES: 503  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/905,223  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelson, Ned A.  
; REGISTRATION NUMBER: 29,655  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: Other nucleic acid  
; US-08-905-223-2  
  
Query Match 100.0%; Score 24; DB 4; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 gcauccuacuccaaucca 24  
DB 1 GCATCCUACUCCCAUCCA 24

```

; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: Other nucleic acid
; FEATURE:
; NAME/KEY: Cap
; LOCATION: 1
; OTHER INFORMATION: m7Gppp added to 1
;
US-08-905-223-1

Query Match          100.0%; Score 24; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccaauucca 24
   ||||||||||||||||||||
Db 2 GCAUCCUACUCCCAUCCAUCCA 25

RESULT 4
US-09-247-155-1
; Sequence 1, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 47
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: in vitro transcription product
;
FEATURE:
; NAME/KEY: modified_base
; LOCATION: 1
; OTHER INFORMATION: m7g
;
US-09-247-155-1

Query Match          100.0%; Score 24; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacucccaauucca 24
   ||||||||||||||||||||
Db 2 gcauccuacucccaauucca 25

RESULT 5
US-08-930-102A-1
; Sequence 1, Application US/08930102A
; Patent No. 6022715

```

GENERAL INFORMATION:  
APPLICANT: Dumas, Jean-Baptiste Milne Edwards  
APPLICANT: Merenkova, Irena Nikolaevna  
TITLE OF INVENTION: METHOD FOR THE SPECIFIC COUPLING OF THE CAP  
TITLE OF INVENTION: OF THE 5' END OF AN mRNA FRAGMENT AND PREPARATION OF mRNA AND OF  
TITLE OF INVENTION: COMPLETE CDNA  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930.102A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR96/00651  
FILING DATE: 29-APR-1996  
APPLICATION NUMBER: FR95/05221  
FILING DATE: 02-MAY-1995  
APPLICATION NUMBER: FR95/09467  
FILING DATE: 03-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ned A. Israelsen  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: GENSET .017APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-930-102A-1

Query Match 100.0%; Score 24; DB 3; Length 65;  
Best local Similarity 75.0%; Pred. No. 0.0043;  
Matches 18; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcauccuacucccauccaaucca 24  
|||||:|||||:|||||:|||||  
DB 20 GCATCCCTACTCCCATCCCAATCCCA 43

RESULT 6  
US-08-930-102A-4  
Sequence 4, Application US/08930102A  
Patent No. 6022715  
GENERAL INFORMATION:  
APPLICANT: Dumas, Jean-Baptiste Milne Edwards  
APPLICANT: Merenkova, Irena Nikolaevna  
TITLE OF INVENTION: METHOD FOR THE SPECIFIC COUPLING OF THE CAP  
TITLE OF INVENTION: OF THE 5' END OF AN mRNA FRAGMENT AND PREPARATION OF mRNA AND OF  
TITLE OF INVENTION: COMPLETE CDNA  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, 16th Floor  
CITY: Newport Beach  
STATE: CA

COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930.102A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR96/00651  
FILING DATE: 29-APR-1996  
APPLICATION NUMBER: FR95/05221  
FILING DATE: 02-MAY-1995  
APPLICATION NUMBER: FR95/09467  
FILING DATE: 03-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ned A. Israelsen  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: GENSET .017APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA (genomic)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1  
OTHER INFORMATION: /label=m7Gppp  
OTHER INFORMATION: /note="N is m7Gpppg"  
US-08-930-102A-4

Query Match 95.8%; Score 23; DB 3; Length 46;  
Best local Similarity 100.0%; Pred. No. 0.012;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cauccuacucccauccaaucca 24  
|||||:|||||:|||||:|||||  
DB 2 CAUCCUACUCCCAUCCCAUCCCA 24

RESULT 7  
US-08-930-102A-5  
Sequence 5, Application US/08930102A  
Patent No. 6022715  
GENERAL INFORMATION:  
APPLICANT: Dumas, Jean-Baptiste Milne Edwards  
APPLICANT: Merenkova, Irena Nikolaevna  
TITLE OF INVENTION: METHOD FOR THE SPECIFIC COUPLING OF THE CAP  
TITLE OF INVENTION: OF THE 5' END OF AN mRNA FRAGMENT AND PREPARATION OF mRNA AND  
TITLE OF INVENTION: COMPLETE CDNA  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

Qy 2 cauccuacucccauccaaucca 24  
| :||:| | | | | :|||  
Db 347 CTTCTACCCCCATCCAAC TCCA 325

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/301,162  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,128  
FILING DATE:  
APPLICATION NUMBER: US 07/623,086  
FILING DATE: 06-DEC-1990  
APPLICATION NUMBER: DE P3940598.2  
FILING DATE: 08-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Fleshner, Raz E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 02481.1005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-301-162-13

Query Match 67.5%; Score 16.2; DB 3; Length 250;  
Best Local Similarity 57.1%; Pred. No. 29;  
Matches 12; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 2 cauccuacucccauccaauuc 22  
| :||:|:||||:|:|:|  
Db 178 CTCCTTCTCCCATCGATTC 198

RESULT 12  
US-09-461-240-13  
Sequence 13, Application US/09461240  
Patent No. 6326008  
GENERAL INFORMATION:  
APPLICANT: Knapp, Stefan  
Ziegelmaier, Robert  
Kupper, Hans  
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The Preparation Thereof and the Use Thereof  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/461,240  
FILING DATE: 16-DEC-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/301,162  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US/08/167,128  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/623,086  
FILING DATE: 06-DEC-1990  
APPLICATION NUMBER: DE P3940598.2  
FILING DATE: 08-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Fleshner, Raz E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 02481.1005-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-461-240-13

Query Match 67.5%; Score 16.2; DB 4; Length 250;  
Best Local Similarity 57.1%; Pred. No. 29;  
Matches 12; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 2 cauccuacucccauccaauuc 22  
| :||:|:||||:|:|:|  
Db 178 CTCCTTCTCCCATCGATTC 198

RESULT 13  
US-08-170-294-5  
Sequence 5, Application US/08170294  
Patent No. 5389614  
GENERAL INFORMATION:  
APPLICANT: BRIDGES, IAN G.  
APPLICANT: BRIGHT, SIMON W.J.  
APPLICANT: GREENLAND, ANDREW J.  
APPLICANT: HOLT, DAVID C.  
APPLICANT: JEPSON, IAN  
APPLICANT: SCHUCH, WOLFGANG W.  
TITLE OF INVENTION: PLANT-DERIVED ENZYME AND DNA SEQUENCES  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170,294  
FILING DATE: 30-DEC-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/GB92/01187  
FILING DATE: 01-JUL-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9114259.6  
FILING DATE: 02-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 204218/SEE36438/UST  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: PRIMER EXTENSION MAP FIGURE 5  
US-08-170-294-5

Query Match 66.7%; Score 16; DB 1; Length 305;  
Best Local Similarity 66.7%; Pred. No. 38;  
Matches 16; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauucca 24  
Db 167 GCTGCACACACCCATCCATTCACA 190

RESULT 14  
US-08-664-855-5  
Sequence 5, Application US/08664855

Patent No. 5866792

GENERAL INFORMATION:

APPLICANT: BRIDGES, IAN G.

APPLICANT: BRIGHT, SIMON W.J.

APPLICANT: GREENLAND, ANDREW J.

APPLICANT: HOLT, DAVID C.

APPLICANT: JEPSON, IAN

APPLICANT: SCHUCH, WOLFGANG W.

TITLE OF INVENTION: PLANT-DERIVED ENZYME AND DNA SEQUENCES

NUMBER OF SEQUENCES: 7

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/664,855

FILING DATE: 17-JUN-1996

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/170,294

FILING DATE: 30-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9114259.6

FILING DATE: 02-JUL-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB92/01187

FILING DATE: 01-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: KOKULIS, PAUL N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 224452/SEE36438USTD1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 305 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: PRIMER EXTENSION MAP FIGURE 5  
US-08-664-855-5

Query Match 66.7%; Score 16; DB 2; Length 305;  
Best Local Similarity 66.7%; Pred. No. 38;  
Matches 16; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcauccuacucccauucca 24  
Db 167 GCTGCACACACCCATCCATTCACA 190

RESULT 15  
US-09-049-289-5  
Sequence 5, Application US/09049289

Patent No. 6066456

GENERAL INFORMATION:

APPLICANT: BRIDGES, IAN G.

APPLICANT: BRIGHT, SIMON W.J.

APPLICANT: GREENLAND, ANDREW J.

APPLICANT: HOLT, DAVID C.

APPLICANT: JEPSON, IAN

APPLICANT: SCHUCH, WOLFGANG W.

TITLE OF INVENTION: PLANT-DERIVED ENZYME AND DNA SEQUENCES

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DABRY & CUSHMAN L.L.P.

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/049,289

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/170,294

FILING DATE: 30-DEC-1993

APPLICATION NUMBER: WO PCT/GB92/01187

FILING DATE: 01-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9114259.6

FILING DATE: 02-JUL-1991

ATTORNEY/AGENT INFORMATION:

NAME: KOKULIS, PAUL N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 204218/SEE36438/UST

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 305 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: PRIMER EXTENSION MAP FIGURE 5
US-09-049-289-5

Query Match      66.7%; Score 16; DB 3; Length 305;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 16; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcauccuacucccaaucca 24
   ||| | |||:||||:||||
Db 167 GCTGCACACACCCCAATTC 190

Search completed: June 2, 2002, 18:58:39
Job time: 11064 sec
```

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 16:03:15 ; Search time 523.16 Seconds  
(without alignments)  
78.764 Million cell updates/sec

Title: US-09-663-600A-2\_COPY\_1\_24  
Perfect score: 24  
Sequence: 1 gcauccuacuccaauucca 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1980.DAT.\*  
2: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1981.DAT.\*  
3: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1982.DAT.\*  
4: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1983.DAT.\*  
5: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1984.DAT.\*  
6: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1985.DAT.\*  
7: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1986.DAT.\*  
8: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1987.DAT.\*  
9: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1988.DAT.\*  
10: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1989.DAT.\*  
11: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1990.DAT.\*  
12: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1991.DAT.\*  
13: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1992.DAT.\*  
14: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1993.DAT.\*  
15: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1994.DAT.\*  
16: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1995.DAT.\*  
17: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1996.DAT.\*  
18: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1997.DAT.\*  
19: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1998.DAT.\*  
20: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1999.DAT.\*  
21: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2000.DAT.\*  
22: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2001A.DAT.\*  
23: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2001B.DAT.\*  
24: /net/abss06/SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	46	20	AAZ40750
2	24	100.0	46	20	AAZ40750
3	24	100.0	46	20	AAZ40750
4	24	100.0	46	20	AAZ40750
5	24	100.0	46	20	AAZ40750
6	24	100.0	46	20	AAZ40750
7	24	100.0	46	20	AAZ40750
8	24	100.0	46	20	AAZ40750
9	24	100.0	46	20	AAZ40750

10	24	100.0	46	20	AAZ41349
11	24	100.0	46	20	AAZ30063
12	24	100.0	47	17	AAT43581
13	24	100.0	47	20	AAZ40749
14	24	100.0	47	20	AAZ88170
15	24	100.0	47	20	AAZ97543
16	24	100.0	47	20	AAZ19962
17	24	100.0	47	20	AAZ51428
18	24	100.0	47	20	AAZ39409
19	24	100.0	47	20	AAZ40407
20	24	100.0	47	20	AAZ51756
21	24	100.0	47	20	AAZ26651
22	24	100.0	47	20	AAZ41348
23	24	100.0	47	20	AAZ30062
c 24	18.2	75.8	1312	20	AAV64413
c 25	18.2	75.8	6991	20	AAV64412
c 26	18.2	75.8	8176	20	AAV64411
c 27	17.8	74.2	1161	21	AAZ49972
c 28	17.8	74.2	1542	21	AAZ41500
c 29	17.8	74.2	1605	21	AAZ43079
c 30	17.8	74.2	1650	21	AAZ49646
c 31	17.6	73.3	417	21	AAZ06005
c 32	17.6	73.3	3255	23	AAZ68839
c 33	17.2	71.7	633	22	AAH05985
c 34	17.2	71.7	879	22	AAH63420
c 35	17.2	71.7	2119	22	AAH17697
c 36	17.2	71.7	5672	22	AAH15193
c 37	17.2	71.7	6802	22	AAZ46282
c 38	17.2	71.7	6802	24	AAZ61068
c 39	17.2	71.7	34980	22	AAH68525
c 40	16.8	70.0	322	20	AAV86922
c 41	16.8	70.0	329	21	AAZ26424
c 42	16.8	70.0	3797	23	ABL21119
c 43	16.8	70.0	3858	23	ABL11191
c 44	16.8	70.0	4087	22	AAZ27712
c 45	16.8	70.0	6461	23	ABL11190

ALIGNMENTS

RESULT 1  
AAZ40750  
ID AAZ40750 standard; RNA; 46 BP.  
XX  
AC AAZ40750;  
XX  
DT 18-JAN-2000 (first entry)  
XX  
XX Oligonucleotide -Cap for Secreted protein EST isolation.  
XX  
DE  
XX  
XX PCR primer; secreted protein; fingerprint identification technique;  
XX Chromosome mapping; human; hereditary disease; diagnosis; cancer;  
XX hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;  
XX autoimmune disease; rheumatic disease; embryonic disorder; myopathy;  
XX renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
XX hypertension; ss.  
XX  
XX Synthetic.  
XX Homo sapiens.  
XX  
XX WO9940189-A2.  
XX  
XX 12-AUG-1999.  
PD  
XX  
XX 09-FEB-1999; 99WO-IB00282.  
XX  
XX  
XX 09-FEB-1998; 98US-0074121.  
XX 13-APR-1998; 98US-0081563.  
PR 10-AUG-1998; 98US-0096116.  
PR 04-SEP-1998; 98US-0099273.  
XX  
XX (GSET ) GENSET.

```

XX  Bougueleret L, Duclert A, Dumas Milne Edwards J;
PI  WPI: 1999-600966/51.
XX
XX  Extended cDNAs useful for expressing secreted proteins and to obtain
PT  specific antibodies -
XX
PS  Example 2; Page 12; 244pp; English.
XX
XX  This sequence represents a PCR primer used within the course of the
CC  invention. The invention relates to 70 nucleic acids encoding human
CC  secreted proteins. The extended cDNAs (or genomic DNAs obtainable from
CC  them) may be used to prepare PCR primers and probes. These are useful for
CC  forensic matching or positive identification by DNA sequencing. They may
CC  also be used in alternative fingerprint identification techniques.
CC  Antibodies against the proteins encoded by the extended cDNAs are useful
CC  in identification of tissue types or cell species, as well as identifying
CC  tissue specific soluble proteins. The sequences can be used for
CC  chromosome mapping and identification of genes associated with hereditary
CC  diseases or drug response. Signal sequences from the cDNAs can be used in
CC  construction of secretion vectors. Other sequences derived from the
CC  extended cDNAs can be used to clone upstream genomic DNA sequences
CC  including promoters. This is in turn useful for identifying proteins that
CC  interact with promoter sequences. Some of the proteins may be useful in
CC  diagnosing and treating several disorders including, but not limited to:
CC  cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders,
CC  autoimmune diseases, and rheumatic diseases, embryogenic disorders,
CC  hypertension, renal injury, amino acidurias, hypoglycaemia, male rat
CC  infertility and myopathies.
XX
SQ  Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match          100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. NO. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 gcauccuacuccaauuca 24
    ||||||||||||||||
DB  1 gcauccuacuccaauuca 24

RESULT  2
AA88171 standard; RNA; 46 BP.
XX
AC  AAX88171;
XX
DT  23-SEP-1999 (first entry)
XX
DE  Oligoribonucleotide Cap5'-ppp.
XX
KW  Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;
KM  diagnostic; gene therapy; chromosome mapping; secretion vector; primer;
KM  ss.
XX
OS  Synthetic.
XX
PN  WO9925825-A2.
XX
PD  27-MAY-1999.
XX
PF  13-NOV-1998; 98WO-IB01862.
XX
PR  04-SEP-1998; 98US-0099273.
PR  13-NOV-1997; 97US-006677.
PR  17-DEC-1997; 97US-0069957.
PR  09-FEB-1998; 98US-0074121.
PR  13-APR-1998; 98US-0081563.
PR  10-AUG-1998; 98US-0096116.
XX
XX  (GEST ) GENSET.

```

```

XX  Bougueleret L, Duclert A, Dumas Milne Edwards J;
PI  WPI: 1999-347472/29.
XX
XX  Extended cDNAs encoding secreted proteins
PT
XX
PS  Example 2; Page 131; 307pp; English.
XX
XX  This invention describes novel nucleic acid sequences of extended cDNAs
CC  (see AAX97813-X97906) which encode human secreted proteins (see
CC  AAY36129-Y36222) and which have cytosolic, thrombotic and osteopathic
CC  activity. The extended cDNAs can be used to express secreted proteins
CC  or parts of them or to obtain antibodies capable of binding to the
CC  secreted proteins. They may also be used in diagnostic, forensic,
CC  gene therapy and chromosome mapping procedures. Uses also include design
CC  of expression vectors and secretion vectors. This sequence represents
XX  an oligoribonucleotide primer used in the method of the invention.
XX
SQ  Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match          100.0%; Score 24; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. NO. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 gcauccuacuccaauuca 24
    ||||||||||||||||
DB  1 gcauccuacuccaauuca 24

RESULT  3
AAX97544
ID  AAX97544 standard; RNA; 46 BP.
XX
AC  AAX97544;
XX
DT  13-SEP-1999 (first entry)
XX
DE  Oligonucleotide Cap- for secreted protein coding sequence isolation.
XX
KW  Secreted protein; human; cytokine; cellular proliferation; cell movement;
KM  cellular differentiation; immune system regulator; anti-inflammatory;
KM  haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KM  reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KM  genetic disease; ss.
XX
OS  Synthetic.
XX
OS  Homo sapiens.
XX
PN  WO9931236-A2.
XX
PD  24-JUN-1999.
XX
PF  17-DEC-1998; 98WO-IB02122.
XX
PR  10-AUG-1998; 98US-0096116.
PR  17-DEC-1997; 97US-0069957.
PR  09-FEB-1998; 98US-0074121.
PR  13-APR-1998; 98US-0081563.
XX
XX  (GEST ) GENSET.
XX
PI  Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR  WPI: 1999-385906/32.
XX
XX  New isolated human secreted proteins
PT
XX
PS  Example 2; Page 12; 51pp; English.
XX
XX  This sequence represents an oligonucleotide used to isolate the extended
CC  human secreted protein coding sequences of the invention. The secreted

```

CC proteins can be used in treating or controlling a variety of human  
 CC conditions. The secreted proteins may act as cytokines or may affect  
 CC cellular proliferation or differentiation or may act as immune system  
 CC regulators, haematopoiesis regulators, tissue growth regulators,  
 CC regulators of reproductive hormones or cell movement or have  
 CC chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour  
 CC inhibition activity. The DNAs can be used in forensic procedures to  
 CC identify individuals or in diagnostic procedures to identify individuals  
 CC having genetic diseases resulting from abnormal expression of the genes  
 CC corresponding to the extended cDNAs. They are also useful for  
 CC constructing a high resolution map of the human chromosomes. They can  
 CC also be used for gene therapy to control or treat genetic diseases.  
 XX  
 SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcauccuacucccauucca 24  
 |||||  
 Db 1 gcauccuacucccauucca 24

RESULT 4  
 AAX19963  
 ID AAX19963 standard; RNA; 46 BP.  
 AC AAX19963;  
 XX  
 DT 16-JUN-1999 (first entry)  
 DE  
 DE Oligoribonucleotide -Cap SEQ ID NO:2.  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9906439-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01233.  
 XX  
 PR 01-AUG-1997; 97US-0904468.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI; 1999-153700/13.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from liver, lung, large intestine, colon,  
 PT thyroid and pancreas tissue  
 XX  
 PS Example 2; Page 15; 398pp; English.  
 XX  
 CC AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY11533 to  
 CC AAY11679, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductve hormone

CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell. The  
 CC present sequence represents an oligoribonucleotide used in an example  
 CC from the present invention.  
 XX  
 SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcauccuacucccauucca 24  
 |||||  
 Db 1 gcauccuacucccauucca 24

RESULT 5  
 AAX51429  
 ID AAX51429 standard; RNA; 46 BP.  
 XX  
 AC AAX51429;  
 XX  
 DT 21-JUN-1999 (first entry)  
 DE  
 DE Oligonucleotide -Cap.  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9906549-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01231.  
 XX  
 PR 01-AUG-1997; 97US-0905279.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI; 1999-153779/13.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from testis, ovary, uterus and spleen tissue  
 XX  
 PS Example 2; Page 144; 522pp; English.  
 XX  
 CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12681 to  
 CC AAY12913, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductve hormone  
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products

CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide into a cell. This  
CC polypeptide into a membrane, or importing a polypeptide into a cell. This  
CC sequence represents an oligonucleotide -cap, used in the isolation of the  
CC 5' EST sequences of the invention.  
XX  
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;  
  
Query Match 100.0%; Score 24; DB 20; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 gcauccuacucccauccaaucca 24  
|||||  
Db 1 gcauccuacucccauccaaucca 24  
  
RESULT 6  
AAK39410  
ID AAK39410 standard; RNA; 46 BP.  
XX  
AC AAK39410;  
XX  
DT 21-JUN-1999 (first entry)  
XX  
DE Human secreted protein 5' EST detecting oligoribonucleotide #2.  
XX  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
XX  
OS Synthetic.  
XX  
PN WO9906551-A2.  
XX  
PD 11-FEB-1999.  
XX  
PF 31-JUL-1998; 98WO-IB01235.  
XX  
PR 01-AUG-1997; 97US-0905133.  
XX  
PA (GEST ) GENSET.  
XX  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
XX  
DR WPI; 1999-153781/13.  
XX  
PT New nucleic acids encoding human secreted - proteins obtained from  
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals  
PT and fetal brain tissue  
XX  
PS Example 2; Page 15; 434pp; English.  
XX  
CC This invention describes 5' expressed sequence tags (ESTs) represented  
CC in AAK39440 to AAK39597, which encode the human secreted proteins  
CC represented in AAK11374-Y11531. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.  
CC They can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter

CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell. This  
CC sequence represents an oligoribonucleotide which is used in the method of  
CC the invention.  
XX  
SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;  
  
Query Match 100.0%; Score 24; DB 20; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 gcauccuacucccauccaaucca 24  
|||||  
Db 1 gcauccuacucccauccaaucca 24  
  
RESULT 7  
AAK40408  
ID AAK40408 standard; RNA; 46 BP.  
XX  
AC AAK40408;  
XX  
DT 18-JUN-1999 (first entry)  
XX  
DE Oligoribonucleotide -Cap.  
XX  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
XX  
OS Synthetic.  
XX  
PN WO9906550-A2.  
XX  
PD 11-FEB-1999.  
XX  
PF 31-JUL-1998; 98WO-IB01232.  
XX  
PR 01-AUG-1997; 97US-0905144.  
XX  
PA (GEST ) GENSET.  
XX  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
XX  
DR WPI; 1999-153780/13.  
XX  
PT New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
XX  
PS Example 2; Page 15; 675pp; English.  
XX  
CC AAK40438 to AAK40715 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins expressed in prostate, and encode the proteins  
CC given in AAK11716 to AAK11993 respectively. The proteins given represent  
CC the signal peptide and an N-terminal fragment of a secreted protein. The  
CC nucleic acid sequences can be used for producing secreted human gene  
CC products. They can also be used to develop products for diagnosis and  
CC therapy. The proteins obtained may have cytokine activity, cell  
CC proliferation and differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptides can be used for

CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX  
 SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;  
 Query Match 100.0%; Score 24; DB 20; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.13; Mismatches 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gcauccuacucccauucca 24  
 |||||  
 Db 1 gcauccuacucccauucca 24  
 RESULT 8  
 AAX51757  
 ID AAX51757 standard; RNA; 46 BP.  
 XX  
 AC AAX51757;  
 XX  
 DT 22-JUN-1999 (first entry)  
 XX  
 DE Uncapped mRNA for EST sequence of human secreted protein.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906552-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01236.  
 XX  
 PR 01-AUG-1997; 97US-0905223.  
 XX  
 XX (GEST ) GENSET.  
 PA  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI; 1999-153782/13.  
 XX  
 PT New isolated brain-derived nucleic acids - used to develop products  
 PT which may have cytokine, immune, regulatory, haematopoiesis  
 PT regulating, anti-inflammatory or tumour inhibition activity  
 XX  
 PS Example 2; Page 15; 577pp; English.  
 XX  
 CC AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12987 to  
 CC AAY13219, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC This sequence was used in a method to isolate the 5' ESTs of the genes  
 CC encoding the human secreted proteins of the invention.

XX  
 SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;  
 Query Match 100.0%; Score 24; DB 20; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.13; Mismatches 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gcauccuacucccauucca 24  
 |||||  
 Db 1 gcauccuacucccauucca 24  
 RESULT 9  
 AAX26652  
 ID AAX26652 standard; RNA; 46 BP.  
 XX  
 AC AAX26652;  
 XX  
 DT 18-JUN-1999 (first entry)  
 XX  
 DE Oligoribonucleotide used to identify 5' EST sequences.  
 XX  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9906554-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-IB01238.  
 XX  
 PR 01-AUG-1997; 97US-0905134.  
 XX  
 XX (GEST ) GENSET.  
 PA  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI; 1999-153784/13.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from kidney, fetal kidney, dystrophic  
 PT muscle, muscle and heart tissue  
 XX  
 PS Example 2; Page 15; 622pp; English.  
 XX  
 CC The specification describes 5' expressed sequence tags (ESTs), see  
 CC AAX40826-X41093) for human secreted proteins (see AAY01602 and AAY11994-  
 CC AAY12260). The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell. The present sequence is used  
 CC in the course of the invention.

XX  
 SQ Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcauccuacuccaauuca 24  
1 gcauccuacuccaauuca 24  
DB 1 gcauccuacuccaauuca 24

## RESULT 10

AAK41349  
ID AAK41349 standard; RNA; 46 BP.

AC AAK41349;

DT 22-JUN-1999 (first entry)

DE Oligoribonucleotide, SEQ ID NO: 2 from M0906553.

Human; secreted protein; EST; expressed sequence tag; diagnosis;  
forensic; gene therapy; chromosome mapping; signal peptide;  
upstream regulatory sequence; cytokine activity; cell proliferation;  
differentiation; haematopoiesis regulation; tissue growth regulation;  
reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
thrombolytic; antiinflammatory; tumour inhibition; antitumour; ss.

OS Homo sapiens.

PN M0906553-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-1B01237.

PR 01-AUG-1997; 97US-0905051.

PA (GEST ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI; 1999-153783/13.

PT New nucleic acids encoding human secreted proteins - obtained from  
CDNA libraries derived from umbilical cord, lymph ganglia,  
lymphocytes and placental tissue

PS Example 2; Page 15; 411pp; English.

CC The patent relates to sequences of 5' ESTs derived from mRNAs  
CC encoding secreted proteins. The nucleic acid sequences can  
CC be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, antiinflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.

SO Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcauccuacuccaauuca 24

DB 1 gcauccuacuccaauuca 24  
|||||

## RESULT 11

AAK30063  
ID AAK30063 standard; RNA; 46 BP.

AC AAK30063;

DT 17-JUN-1999 (first entry)

DE Oligoribonucleotide -Cap SEQ ID NO:2.

Human; secreted protein; EST; expressed sequence tag; diagnosis;  
forensic; gene therapy; chromosome mapping; signal peptide;  
upstream regulatory sequence; cytokine activity; cell proliferation;  
differentiation; haematopoiesis regulation; tissue growth regulation;  
reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
thrombolytic; anti-inflammatory; tumour inhibition; ss.

OS Synthetic.

PN M0906548-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-1B01222.

PR 01-AUG-1997; 97US-0905135.

PA (GEST ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI; 1999-153778/13.

PT New nucleic acids encoding human secreted proteins - obtained from  
CDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
kidney, lung, umbilical cord, placenta and colon tissue

PS Example 2; Page 15; 824pp; English.

CC AAK41094 to AAK41347 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAY12261 to  
CC AAY12314, respectively. The proteins given represent the signal peptide  
CC and an N-terminal fragment of a secreted protein. The nucleic acid  
CC sequences can be used for producing secreted human gene products. They  
CC can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell. The  
CC present sequence represents an oligoribonucleotide used in an example  
CC from the present invention.

SO Sequence 46 BP; 10 A; 24 C; 1 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcauccuacuccaauuca 24  
1 gcauccuacuccaauuca 24  
DB 1 gcauccuacuccaauuca 24

```
RESULT 12
AAT43581
ID AAT43581 standard; mRNA; 47 BP.
XX
AC AAT43581;
XX
DT 04-AUG-1997 (first entry)
XX
DE 5'-capped mRNA for coupling to blotin label via amino linker.
XX
KW Messenger RNA; guanosine 5'-cap; label: immobilisation; capture;
KW polymerase chain reaction; transcription template; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT modified_base 1 /*tag= a
FT /mod_base= m7g
FT /function= cap
FT /note= "linked to adjacent nucleotide via
FT (5'ppp(5') linkage"
XX
XX WO9634981-A2.
XX
XX 07-NOV-1996.
XX
XX 29-APR-1996; 96WO-FR00651.
XX
XX 03-AUG-1995; 95FR-0009467.
XX
XX 02-MAY-1995; 95FR-0005221.
XX
XX (GENQ ) GENSET.
XX
XX Dumas Milne Edwards JG, Nicolaevna Merenkova I;
XX
XX WPI; 1996-506181/50.
XX
XX Specific coupling of the 5' cap of mRNA to amino-functionalised cpd.
XX -by eliminating 3' diol, oxidn. Of cap diol to aldehyde and
XX reaction with amine, e.g. for isolation of complete RNA, labelling
XX etc.
XX
XX Example 1; Page 20; 49pp; French.
XX
XX The first step in a new method for specifically coupling the cap of
XX the 5'-end of a eukaryotic mRNA to an amino-functionalised compound
XX involves specifically modifying the 3'-end of the mRNA so that the
XX last base no longer contains OH groups at the 2' and 3' positions.
XX Then, the 2', 3'-cis diol of the methyl guanosine residue at the 5'-end
XX can be oxidised to form a 2', 3'-dialdehyde which is ultimately coupled
XX with the amino group of the amino-functionalised compound. The method
XX is used to label specifically at the 5'-cap, to isolate the 5'-end
XX of mRNA in a sample, to produce the 3'-end of cDNA, to produce double
XX stranded cDNA complementary to the 5'-end of mRNA or to capture mRNA-
XX binding proteins. In a specific example of the coupling method, a
XX double-stranded template was prepared by PCR amplification using a
XX 5'-primer containing the T7 RNA polymerase promoter (AAT43579) and
XX a 3'-primer (AAT43580). When the template was transcribed in the
XX presence of cap analogue m7G(5')ppp(5')G, a capped RNA transcript
XX having the present sequence was produced. The 5'-capped mRNA was
XX coupled to a blotin label via a hydrazine linker.
XX
XX Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;
XX
XX Query Match 100.0%; Score 24; DB 17; Length 47;
XX Best Local Similarity 100.0%; Pred. No. 0.13;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 gcauccuacucccauccaaucca 24
XX
XX Db 2 gcauccuacucccauccaaucca 25
XX
XX RESULT 13
XX AAZ40749
XX ID AAZ40749 standard; RNA; 47 BP.
XX
XX AC AAZ40749;
XX
XX DT 18-JAN-2000 (first entry)
XX
XX DE Oligonucleotide 5'-cap for Secreted protein EST isolation.
XX
XX KW PCR primer; secreted protein; fingerprint identification technique;
XX chromosome mapping; human; hereditary disease; diagnosis; cancer;
XX hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
XX autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
XX renal injury; amino aciduria; hypoglycaemia; male rat infertility;
XX hypertension; ss.
XX
XX OS Synthetic.
XX
XX OS Homo sapiens.
XX
XX XX WO9940189-A2.
XX
XX PD 12-AUG-1999.
XX
XX PF 09-FEB-1999; 99WO-IB00282.
XX
XX XX 09-FEB-1998; 98US-0074121.
XX
XX PR 13-APR-1998; 98US-0081563.
XX
XX PR 10-AUG-1998; 98US-0096116.
XX
XX PR 04-SEP-1998; 98US-0099273.
XX
XX XX (GEST ) GENSET.
XX
XX PA Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
XX PI WPI; 1999-600966/51.
XX
XX DR Extended cDNAs useful for expressing secreted proteins and to obtain
XX specific antibodies -
XX
XX PS Example 2; Page 12; 244pp; English.
XX
XX CC This sequence represents a PCR primer used within the course of the
XX invention. The invention relates to 70 nucleic acids encoding human
XX secreted proteins. The extended cDNAs (or genomic DNAs obtainable from
XX them) may be used to prepare PCR primers and probes. These are useful for
XX forensic matching or positive identification by DNA sequencing. They may
XX also be used in alternative fingerprint identification techniques. They may
XX antibodies against the proteins encoded by the extended cDNAs are useful
XX in identification of tissue types or cell species, as well as identifying
XX tissue specific soluble proteins. The sequences can be used for
XX chromosome mapping and identification of genes associated with hereditary
XX diseases or drug response. Signal sequences from the cDNAs can be used in
XX construction of secretion vectors. Other sequences derived from the
XX extended cDNAs can be used to clone upstream genomic DNA sequences
XX including promoters. This is in turn useful for identifying proteins that
XX interact with promoter sequences. Some of the proteins may be useful in
XX diagnosing and treating several disorders including, but not limited to:
XX cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders,
XX autoimmune diseases, and rheumatic diseases, embryogenic disorders,
XX hypertension, renal injury, amino acidurias, hypoglycaemia, male rat-
XX infertility and myopathies.
XX
XX SQ Sequence 47 BP; 10 A; 24 C; 1 G; 11 U; 1 other;
XX
XX Query Match 100.0%; Score 24; DB 20; Length 47;
XX Best Local Similarity 100.0%; Pred. No. 0.13;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 gcauccuacuccaaucca 24  
 XX ||||||||||||||||||  
 DB 2 gcauccuacuccaaucca 25

RESULT 14  
 AAX88170  
 ID AAX88170 standard; RNA; 47 BP.  
 XX  
 AC AAX88170;  
 XX  
 DF 23-SEP-1999 (first entry)  
 XX

DE Oligoribonucleotide 5'm/Gppp.

XX Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;  
 KW diagnostic; gene therapy; chromosome mapping; secretion vector; primer;  
 KW ss.  
 XX

OS Synthetic.

PN W09925825-A2.

PD 27-MAY-1999.

PF 13-NOV-1998; 98WO-1B01862.

PR 04-SEP-1998; 98US-0099273.

PR 13-NOV-1997; 97US-0066677.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PA (GEST ) GENSET.

PI Bouguerel L, Duclert A, Dumas Milne Edwards J;

PS WPI; 1999-347472/29.

PT Extended cDNAs encoding secreted proteins

XX Example 2; Page 131; 307pp; English.

CC This invention describes novel nucleic acid sequences of extended cDNAs  
 CC (see AAX97813-X97906) which encode human secreted proteins (see  
 CC AAY6129-Y56222) and which have cytosolic, thrombotic and osteopathic  
 CC activity. The extended cDNAs can be used to express secreted proteins  
 CC or parts of them or to obtain antibodies capable of binding to the  
 CC secreted proteins. They may also be used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. Uses also include design  
 CC of expression vectors and secretion vectors. This sequence represents  
 CC an oligonucleotide primer used in the method of the invention.

CC Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcauccuacuccaaucca 24  
 XX ||||||||||||||||||  
 DB 2 gcauccuacuccaaucca 25

RESULT 15  
 AAX97543  
 ID AAX97543 standard; RNA; 47 BP.  
 XX  
 AC AAX97543;  
 XX

DF 13-SEP-1999 (first entry)  
 XX  
 DE Oligonucleotide Cap+ for secreted protein coding sequence isolation.  
 XX  
 KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
 KW cellular differentiation; immune system regulator; anti-inflammatory;  
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
 KW genetic disease; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT modified\_base 1  
 FT /\*tag= a  
 FT /mod\_base= 7-methylguanine  
 XX

PN W09931236-A2.

PD 24-JUN-1999.

PF 17-DEC-1998; 98WO-1B02122.

PR 10-AUG-1998; 98US-0096116.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PA (GEST ) GENSET.

PI Bouguerel L, Duclert A, Dumas Milne Edwards J;

PS WPI; 1999-385906/32.

PT New isolated human secreted proteins

XX Example 2; Page 12; 516pp; English.

CC This sequence represents an oligonucleotide used to isolate the extended  
 CC human secreted protein coding sequences of the invention. The secreted  
 CC proteins can be used in treating or controlling a variety of human  
 CC conditions. The secreted proteins may act as cytokines or may affect  
 CC cellular proliferation or differentiation or may act as immune system  
 CC regulators, haematopoiesis regulators, tissue growth regulators,  
 CC chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour  
 CC inhibition activity. The DNAs can be used in forensic procedures to  
 CC identify individuals or in diagnostic procedures to identify individuals  
 CC having genetic diseases resulting from abnormal expression of the genes  
 CC corresponding to the extended cDNAs. They are also useful for  
 CC constructing a high resolution map of the human chromosomes. They can  
 CC also be used for gene therapy to control or treat genetic diseases.

CC Sequence 47 BP; 10 A; 24 C; 2 G; 11 U; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcauccuacuccaaucca 24  
 XX ||||||||||||||||||  
 DB 2 gcauccuacuccaaucca 25

Search completed: June 2, 2002, 20:08:48  
 Job time: 14733 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 15:53:00 ; Search time 2378.87 Seconds  
(without alignments)  
211.124 Million cell updates/sec

Title: US-09-663-600A-2\_COPY\_1\_24

Perfect score: 24

Sequence: 1 gcauccuacuccaaucca 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
------------	-------	--------------	----	-------------

1	24	100.0	46	6	ARI77314	ARI77314 Sequence
2	24	100.0	47	6	ARI77313	ARI77313 Sequence
3	24	100.0	65	6	A58296	A58296 Sequence 1
4	23	95.8	46	6	A58299	A58299 Sequence 4
5	23	95.8	46	6	A58300	A58300 Sequence 5
6	19.2	80.0	196356	2	AC096552	AC096552 Homo sapi
7	19.2	80.0	336028	2	AC092096	AC092096 Mus muscu
8	18.8	78.3	47295	8	F22M8	AC020622 Sequence
9	18.8	78.3	78249	9	AP004529	AP004529 Lotus jip
10	18.8	78.3	100849	9	HUNNEUROF	L05367 Human oligo
11	18.8	78.3	142898	2	AC095140	AC095140 Rattus no
12	18.8	78.3	200768	2	AC084272	AC084272 Homo sapi
13	18.8	78.3	203504	2	AC068162	AC068162 Homo sapi
14	18.8	78.3	204991	2	AC068980	AC068980 Homo sapi
15	18.8	78.3	204993	2	AC024447	AC024447 Homo sapi
16	18.8	78.3	205825	2	AC092942	AC092942 Homo sapi
17	18.8	78.3	297898	9	AC004526	AC004526 Homo sapi
18	18.4	76.7	146722	2	AL139254	AL139254 Homo sapi
19	18.4	76.7	245341	2	AC073436	AC073436 Mus muscu
20	18.2	75.8	1312	10	AF016040	AF016040 Mus muscu
21	18.2	75.8	1489	4	AB004272	AB004272 Bos tauru
22	18.2	75.8	6635	8	D14487	D14487 Lentinus ed
23	18.2	75.8	8172	10	AF017112	AF017112 Mus muscu
24	18.2	75.8	73605	9	AL136309	AL136309 Human DNA
25	18.2	75.8	73778	6	AX344575	AX344575 Sequence
26	18.2	75.8	112054	2	AC068730	AC068730 Homo sapi
27	18.2	75.8	115154	2	AP003804	AP003804 Oryza sat
28	18.2	75.8	136299	9	HS359N14	297206 Human DNA s
29	18.2	75.8	153615	2	AL645665	AL645665 Homo sapi
30	18.2	75.8	155523	2	AC025367	AC025367 Homo sapi
31	18.2	75.8	160054	2	AP003179	AP003179 Homo sapi
32	18.2	75.8	167214	2	AC099216	AC099216 Rattus no
33	18.2	75.8	167263	2	AC063966	AC063966 Homo sapi
34	18.2	75.8	169635	2	HSAC001235	AC001235 Homo sapi
35	18.2	75.8	190694	2	AC098056	AC098056 Rattus no
36	18.2	75.8	193029	2	AL355594	AL355594 Homo sapi
37	18.2	75.8	208123	2	AL662894	AL662894 Mus muscu
38	18.2	75.8	349980	6	AX344574	AX344574 Sequence
39	17.8	74.2	1867	3	DROPERIODJ	L07827 Drosophila
40	17.8	74.2	1867	3	DROPERIODM	L07824 Drosophila
41	17.8	74.2	1867	3	DROPERIODO	L07822 Drosophila
42	17.8	74.2	1867	3	DROPERIODQ	L07820 Drosophila
43	17.8	74.2	1870	3	DROPERIODC	L07831 Drosophila
44	17.8	74.2	1870	3	DROPERIODI	L07828 Drosophila
45	17.8	74.2	1870	3	DROPERIODZ	L07811 Drosophila

ALIGNMENTS

RESULT	1					
ARI77314	ARI77314	Sequence 2 from patent US 6312922.	46 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	ARI77314	Sequence 2 from patent US 6312922.				
DEFINITION	ARI77314	Sequence 2 from patent US 6312922.				
ACCESSION	ARI77314	Sequence 2 from patent US 6312922.				
VERSION	ARI77314.1	GI:17919669				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 46)					
AUTHORS	Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.					
TITLE	Complementary DNAs					
JOURNAL	Patent: US 6312922-A 2 06-NOV-2001;					
FEATURES	Location/Qualifiers					
source	1. .46					
BASE COUNT	10 a 24 c 1 g 11 t					
ORIGIN						

Query Match 100.0%; Score 24; DB 6; Length 46;  
Best Local Similarity 75.0%; Pred. No. 0.33;

Matches 18; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacuccaauuca 24  
||||:||||:||||:||||  
Db 1 GCATCCTACTCCATCCATTCCTCA 24

RESULT 2  
LOCUS AR177313 47 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 1 from patent US 6312922.  
ACCESSION AR177313  
VERSION AR177313.1 GI:17919668  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 47)  
AUTHORS Edwards,J.-B.,Dumas,Milne., Duclert,A. and Bougueleret,L.  
TITLE Complementary DNAs  
JOURNAL Patent: US 6312922-A 1 06-NOV-2001;  
FEATURES  
source . 1..47  
/organism="unknown"  
BASE COUNT 10 a 24 c 2 g 11 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 47;  
Best Local Similarity 75.0%; Pred. No. 0.33;  
Matches 18; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacuccaauuca 24  
||||:||||:||||:||||  
Db 2 GCATCCTACTCCATCCATTCCTCA 25

RESULT 3  
LOCUS A58296 65 bp DNA linear PAT 05-MAR-1998  
DEFINITION Sequence 1 from Patent WO9634981.  
ACCESSION A58296  
VERSION A58296.1 GI:3713960  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 65)  
AUTHORS Nicolaevna,M.I. and Dumas,M.E.  
TITLE METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF A FRAGMENT MRNA AND PREPARATION OF MRNA AND COMPLETE CDNA  
JOURNAL Patent: WO 9634981-A 1 07-NOV-1996;  
COMMENT Other publication AU 5982996 961121  
Other publication FR 2733765 961108  
Other publication FR 2733762 961108.  
FEATURES  
source 1..65  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 17 a 29 c 2 g 17 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 65;  
Best Local Similarity 75.0%; Pred. No. 0.33;  
Matches 18; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcauccuacuccaauuca 24  
||||:||||:||||:||||  
Db 20 GCATCCTACTCCATCCATTCCTCA 43

RESULT 4  
LOCUS A58299 46 bp DNA linear PAT 05-MAR-1998  
DEFINITION Sequence 4 from Patent WO9634981.  
ACCESSION A58299  
VERSION A58299.1 GI:3713963  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Nicolaevna,M.I. and Dumas,M.E.  
TITLE METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF A FRAGMENT MRNA AND PREPARATION OF MRNA AND COMPLETE CDNA  
JOURNAL Patent: WO 9634981-A 4 07-NOV-1996;  
COMMENT Other publication AU 5982996 961121  
Other publication FR 2733765 961108  
Other publication FR 2733762 961108.  
FEATURES  
source 1..46  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 10 a 24 c 0 g 11 t 1 others  
ORIGIN

Query Match 95.8%; Score 23; DB 6; Length 46;  
Best Local Similarity 73.9%; Pred. No. 1;  
Matches 17; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cauccuacuccaauuca 24  
|||:||||:||||:||||  
Db 2 CATCCTACTCCATCCATTCCTCA 24

RESULT 5  
LOCUS A58300 46 bp DNA linear PAT 05-MAR-1998  
DEFINITION Sequence 5 from Patent WO9634981.  
ACCESSION A58300  
VERSION A58300.1 GI:3713964  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Nicolaevna,M.I. and Dumas,M.E.  
TITLE METHOD FOR THE SPECIFIC COUPLING OF THE CAP OF THE EXTREMITY 5' OF A FRAGMENT MRNA AND PREPARATION OF MRNA AND COMPLETE CDNA  
JOURNAL Patent: WO 9634981-A 5 07-NOV-1996;  
COMMENT Other publication AU 5982996 961121  
Other publication FR 2733765 961108  
Other publication FR 2733762 961108.  
FEATURES  
source 1..46  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 10 a 24 c 0 g 11 t 1 others  
ORIGIN

Query Match 95.8%; Score 23; DB 6; Length 46;  
Best Local Similarity 73.9%; Pred. No. 1;  
Matches 17; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cauccuacuccaauuca 24  
|||:||||:||||:||||  
Db 2 CATCCTACTCCATCCATTCCTCA 24

RESULT 6

AC096552 196356 bp DNA linear HTG 13-DEC-2001  
Homo sapiens chromosome 2 clone RP11-41M22, WORKING DRAFT SEQUENCE,  
8 unordered pieces.  
ACCESSION AC096552  
VERSION GI:17647039  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 196356)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 196356)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Dec 13, 2001 this sequence version replaced gi:15638726.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: H\_NH0041M22  
Drafting center: WIBR  
-----

----- Summary Statistics -----  
Sequencing vector: M13, 29%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 94% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 193421 bases at least Q40  
Consensus quality: 194158 bases at least Q30  
Consensus quality: 194505 bases at least Q20  
Insert size: 196000; agarose-fp  
Insert size: 195193; sum-of-contigs  
Quality coverage: 12.65 in Q20 bases; agarose-fp  
Quality coverage: 12.67 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1966: contig of 1966 bp in length  
\* 1967 2066: gap of unknown length  
\* 2067 3358: contig of 1292 bp in length  
\* 3359 3458: gap of unknown length  
\* 3459 12020: contig of 8562 bp in length  
\* 12021 12120: gap of unknown length  
\* 12121 32759: contig of 20639 bp in length  
\* 32760 32859: gap of unknown length  
\* 32860 192962: contig of 160103 bp in length  
\* 192963 193063: gap of unknown length  
\* 193064 193255: contig of 463 bp in length  
\* 193256 193625: gap of unknown length  
\* 193626 194887: contig of 1262 bp in length  
\* 194888 194988: gap of unknown length  
\* 194989 196356: contig of 1369 bp in length.

FEATURES  
source  
1. .196356  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="2"  
/clone="RP11-41M22"  
1. .1966  
/note="assembly\_name:Contig10"  
2067. .3358  
/note="assembly\_name:Contig11"  
3459. .12020  
/note="assembly\_name:Contig12"  
clone\_end:T7  
vector\_side:left  
12121. .32759  
/note="assembly\_name:Contig13"  
32860. .192962  
/note="assembly\_name:Contig14"  
193063. .193525  
/note="assembly\_name:Contig3"  
193626. .194887  
/note="assembly\_name:Contig8"  
194988. .196356  
/note="assembly\_name:Contig9"  
BASE COUNT 57531 a 35141 c 38098 g 64879 t 707 others  
ORIGIN

Query Match 80.0%; Score 19.2; DB 2; Length 196356;  
Best Local Similarity 62.5%; Pred.No.64;  
Matches 15; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gcauccuacucccauucca 24  
||||:||||:||||:||||:|  
Db 188328 GCATCTACTACCAATCTTA 188351

RESULT 7  
AC092096/c  
LOCUS AC092096 336028 bp DNA linear HTG 05-FEB-2002  
DEFINITION Mus musculus chromosome 10 clone rp23-39k4, WORKING DRAFT SEQUENCE,  
19 unordered pieces.

ACCESSION AC092096  
VERSION AC092096.17 GI:18497142  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 336028)  
AUTHORS Do,T. and Roe,B.A.  
TITLE Mus musculus Chromosome 10 BAC Clone rp23-39k4  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 336028)  
AUTHORS Do,T. and Roe,B.A.  
TITLE Direct Submission

JOURNAL Submitted (20-JUN-2001) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Feb 5, 2002 this sequence version replaced gi:18201826.

----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 3074: contig of 3074 bp in length  
\* 3075 3174: gap of unknown length  
\* 3175 6014: contig of 2840 bp in length  
\* 6015 6114: gap of unknown length

\* 6115 9345: contig of 3231 bp in length  
\* 9346 12107: contig of 2662 bp in length  
\* 12108 12207: gap of unknown length  
\* 12208 18756: contig of 6549 bp in length  
\* 18757 18857: gap of unknown length  
\* 18857 23733: contig of 4877 bp in length  
\* 23734 23833: gap of unknown length  
\* 23834 30446: contig of 6613 bp in length  
\* 30447 30546: gap of unknown length  
\* 30547 37867: contig of 7321 bp in length  
\* 37868 37968: gap of unknown length  
\* 37968 45708: contig of 7741 bp in length  
\* 45709 53387: gap of unknown length  
\* 53388 53487: gap of unknown length  
\* 53488 63026: contig of 9539 bp in length  
\* 63027 63127: gap of unknown length  
\* 63127 74701: contig of 11574 bp in length  
\* 74701 74801: gap of unknown length  
\* 74801 85536: contig of 10736 bp in length  
\* 85536 85637: gap of unknown length  
\* 85637 108073: contig of 22437 bp in length  
\* 108074 137882: gap of unknown length  
\* 137883 137983: contig of 29709 bp in length  
\* 137983 183434: gap of unknown length  
\* 183434 222828: contig of 45451 bp in length  
\* 222829 222929: gap of unknown length  
\* 222929 262369: gap of unknown length  
\* 262370 336028: contig of 39341 bp in length  
\* 336028 73659: gap of unknown length  
\* 73659 73659: contig of 73659 bp in length.  
Location/Qualifiers  
1. 336028  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/clone="Fp23-39k4"  
/clone.lib="Fp23"

BASE COUNT 101266 a 67078 c 66482 g 99351 t 1851 others  
ORIGIN

Query Match 80.0%; Score 19.2; DB 2; Length 336028;  
Best Local Similarity 66.7%; Pred. No. 63;  
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OR 1 gaucucacuccaauca 24

DB 326247 GCATCTACTCCTCCCAATCCA 326224

RESULT 8  
LOCUS F22M8 47295 bp DNA linear PLN 05-JAN-2001  
DEFINITION Sequence of BAC F22M8 from Arabidopsis thaliana chromosome 1,  
complete sequence.  
ACCESSION AC020622  
VERSION AC020622.3 GI:12039253  
KEYWORDS HTG.  
SOURCE  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 47295)  
Liu, S.X., Sakano, H., Yu, G., Lee, J.M., Lenz, C., Pham, P., Toriumi, M.,  
Chin, C., Chiu, J., Choi, E., Chung, M., Gonzalez, A., Hwang, B.,  
Liu, A., Vaysberg, M., Alfalfi, H., Brooks, S., Buehler, E., Chao, O.,  
Conn, L., Conway, A.B., Hansen, N.F., Johnson-Hopson, C., Khan, S.,  
Kim, C., Lam, B., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P.,  
Southwick, A., Davis, R.W., Ecker, J.R., Federspiel, N.A. and

TITLE Theologis, A.  
JOURNAL The sequence of BAC F22M8 from Arabidopsis thaliana chromosome 1  
Unpublished  
2 (bases 1 to 47295)  
REFERENCE Theologis, A.  
AUTHORS Direct Submission  
JOURNAL Submitted (06-JUN-2000) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
3 (bases 1 to 47295)  
REFERENCE Theologis, A.  
AUTHORS Direct Submission  
JOURNAL Submitted (14-JUN-2000) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
4 (bases 1 to 47295)  
REFERENCE Theologis, A.  
AUTHORS Direct Submission  
JOURNAL Submitted (17-JUN-2000) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
5 (bases 1 to 47295)  
REFERENCE Theologis, A.  
AUTHORS Direct Submission  
JOURNAL Submitted (05-JAN-2001) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA

COMMENT On Jan 5, 2001 this sequence version replaced gi:6693373.  
The sequence is of BAC F22M8 from Arabidopsis thaliana chromosome  
1. In order to facilitate the joining of overlapping clones in the  
future for creation of larger contigs, we provide overlaps between  
overlapping submitted clones. The 3' end of this sequence overlaps  
by 2482 bp the 5' end of the sequence of the clone T123.  
Location/Qualifiers

FEATURES  
source

gene  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/cultivar="Columbia"  
/chromosome="1"  
/clone="F22M8"  
complement(<375..1476)  
/gene="F22M8.1"  
complement(join(<375..459,548..643,748..903,986..1135,  
1246..1476))  
/gene="F22M8.1"

CDS  
/note="Identical to dimethyladenosine transferase (PFC1)  
from Arabidopsis thaliana gb|AF051326 and contains a  
ribosomal RNA Adenine Dimethylases PF100398 domain. This  
gene may be cut off."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76480.1"

trna  
/db\_xref="GI:8570453"  
/translation="MMAVITSAATINCNSLSPSWTCGDNPSKLLLEISMAALRRRT  
VKVSCGSSPDYHSTLKSLSRGRPRKSLGQHYMLNSDINQLASADYKEDFVL  
EIGPRTSLNVLINIGATYALIKDPHMDVLSERFAGSNKFEVLEADPFKIRSH  
MSTLERRLSHPDSALAKVSNLPNISTDVVALLPKMDFSKVILLIODEAALRL  
VERALRTSEYRPINILINFTS"  
1585..1657  
/note="codon recognized: GUA"  
/product="tRNA-Val"  
complement(11759..4192)  
/gene="F22M8.2"

gene  
/note="Contains similarity to excision repair protein  
ERCC5 from Homo sapiens gi11082359 and contains XPG  
N-terminal PF100752 and XPG T-region PF100867 domains."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76467.1"

CDS  
/db\_xref="GI:8570440"  
/translation="MGVGNFMDLRLPRADQGGDFLNKNVANDLSFWIYQHTAYK  
GVLKPHURLFTFTINDLFSKGAIVPFVVDGTPSLKQARISFRFRSSGIDICNLP  
VIRKGVSEVRNKLSEWVRCEVLELIGIVLKNAGAEALCAQLNSOGFVACITP  
DSDAFLEGCAMCIVIKIRPNREPECYHMSHIESGLIKRKLIAISILVNDYDSGG

VLGIGVDKALRIIVREFSEDOVLERLODIGNLQPAVPGGIGKSGDGEFRSEMKRRSP  
HCSRCHLGSKRTHKSCCHCDSGCIKKPLGRCCECSFCSKDRDLREOKKTNW  
IKVCDIALAPEPNRKIIELIYLDGLMTGSGSNMGTPTDGLMYDLWVNDGDCF  
LHDEICIGVQSPADPETERHLEKKURSKKNVSEETATPRATNGVORSITDFY  
RSKAKAAGSIETGGSSKASAEKKRQATSTSSNLTKSRRRLFG  
complement(4476..4548)  
/notes="codon recognized: AAG"  
/product="tRNA-Lys"  
5587..8212  
/gene="F22M8.3"  
join(5587..6182,6538..8212)  
/gene="F22M8.3"  
/notes="Contains similarity to p69d gene from *Lycoopersicon*  
*esculentum* gb|Y17278 and contains a Peptidase 58 Pf|00082  
domain."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76468.1"  
/db\_xref="GI:8570441"  
/translation="MFRSRIVFFLIFFASNVSSRQTVVTHVTSTKHIVTSLFN  
LQTEINDDPSLEPIHYIENAMSGFSATLTDQDQDTVNTKRGFISAPDELLSLH  
TYSHEFLGFEFGIIGWNETSLSDVIIGLVDGTGISEPHVSFRDTHMTPVSRWGGC  
DETFNFSGCKNKKIIGASAFYKYESIVGINKETDTRSTROAGHGHTASTAAGD  
IVPKANYFGAKGLAGMRFTSRIAYKACWALGCASTDVIAADRAILDGVDYISLS  
LGSSRPFXVDPTAAGFGAMOKNIFVSCSAGNSGPTASTVSGCAPLMTVAASYDR  
TPPAIVRIGNRKSIVGSSLYKGSKLNLPLAFNPTAGEAGVPCIRDSLREIVEGR  
IVICLKGASTAGKEEVKRSKGAALLVSTAEGEELLADPHVLPVSLGSDGKTL  
LNYLAGAANATSVFRGTAYGATAPMVAASFSSRGPVAGPEIAKPDIAAGPLNLG  
WSPFSFSLRLSDRRVQVNIISGTSMACPHSIIAALIKSVHGDWSPAMIKSAIMT  
ARTIDNRNPIGRGAAGAESATAFAFGAGNVDPTRAVDPLVDTSTVDYLVNLC  
LANTSERILFESTNTVTCASNAVLSFGDLNYPSEFANLVNGLNLTVRKRYRTVNVG  
SPICEITHVHEEPKRVKRVPEKVLAFQKARERLSYTVTYDAEASRNSSSSFGVLVW  
ICDKYNVRSPVATWE"  
complement(8796..11032)  
/gene="F22M8.4"  
complement(join(8796..8960,9081..9196,9288..9426,  
9666..9752,9853..10003,10092..10216,10778..11032))  
/gene="F22M8.4"  
/notes="Contains similarity to arsenite translocating  
ATPase from *Homo sapiens* gb|AP047469 and contains a 4Fe-4S  
iron sulfur cluster binding protein Pf|00142 domain. EST  
gb|N37510 comes from this gene."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76469.1"  
/db\_xref="GI:8570442"  
/translation="MAADLPEATVQNLQESLKWVFGKGVGKTKCSSILAICLA  
SVRSVLIISTDPAHLSDAFOQFTKSTPLVQESNLFAMEYDPTVETDDMAGTGM  
DGLFSLANAPGIDAMSAEMKLVQTMNDYATVFDPTAGTGLTTPPTALEKG  
LSKMLSKRFGGLMTQMSRMFGMEDEFGEDALIGLEKIDVIEQDMTQFVVCPIPE  
FLSLYETERLVQELARFEIDTHNIIINQVLYDDEDESLLRARMRQOQVLDQFYL  
YDFNTTKLPLLEPVTGVEALKAFSHKFLTPYHPTTSRNVSELEKRVHTLRQLKT  
AEEELERVKSG"  
11405..14708  
/gene="F22M8.5"  
join(11405..11458,11545..11641,11740..11914,11993..12091,  
12368..12394,12495..12595,12754..12843,12966..13046,  
13180..13241,13346..13402,13497..13634,14187..14302,  
14405..14497,14573..14708)  
/gene="F22M8.5"  
/notes="Contains similarity to a hypothetical protein  
g|16983878 from *Oryza Sativa* BAC gb|AP001168."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76470.1"  
/db\_xref="GI:8570443"  
/translation="MAISEEAKLERFLDWLQVNGGELRCNIIKYSLSKFGIFAST  
STQADEVLVPLDIAITPVRVLQDPLIGPEQKMFQGVDDDFLMILFTLERLR  
INSNKPYLDMLPTRCNPLWFSDDDLLELKGNTLHATDKVSFEHFLWANSVPWSRA  
LNLPHPFVFQSDDTGCTSTKIQOAPAPSGSDTIVVEGLVPIDFCNHLKP  
VATWEVDIGSVSRVPSMYLLSVAORIPKKEISISYGNKNEELIYLGFDINDNP  
DYLMIKEMLVNPLVTSVVTFFNNGFIQVHPVPEAIPSPDSKGLLEAQLVESRGG  
EQPSETEVRMAVNEAGDGSALQLLLVLDLLNSKMKMLENSGTEDQARLLEECVLES

HEESRDLDGRRMRNKSWSVVYRRGQKQLTRLLLKAEHALHLSSDH"  
complement(15242..18010)  
/gene="F22M8.6"  
complement(join(15242..15346,15453..15497,15581..15658,  
15970..16623,16952..17066,17159..17607,17714..18010))  
/gene="F22M8.6"  
/notes="Contains similarity to an unnamed protein from *Homo*  
*sapiens* gb|AK001277 and contains an Ank repeat Pf|00023  
domain."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76471.1"  
/db\_xref="GI:8570444"  
/translation="WATTGAAATTEIGNPEYKQPSIFDLTADFDSCSLSNPSETQS  
AFGPKTFDEEEDSKSGVILDRWTCNTKIEFLSQDQRFHFKDIHLNLKIL  
SVQSKAILKEEDVDLTESVDQYDVSISGSEDEAETFPDFHDAQGIDKKILKL  
RLOSQDKVSIWKCLIMDDAESVFNDRGVSVDCCGSLVENEVTERLNLIRENKDDR  
LMRYVLLASGGHFGAGTVNGKSVVAHKTFRHVVRKAKGKOSTKDSGSRSHSAGAS  
QVRYNELALKDIOELLASWKPYFDGAACVFVHAPSSSRLQLLFGKDGKPYFSSQNCAYR  
NVPTIRRTFRFESQRIYNQLTQIAHVTEEIFVNRPEVTKANTVVOVTHNEDSGKTSRK  
EPPDETSSNIIIEEPNRIEEDIEDGTGTSTALHEAAKSGDCERVMFLEEGMDPCA  
KDERGRTPYLANEKEVRNTRFRFALNLEKNWHDAKVPSPLTKEMESQAQAK  
DKOKARTKELKLRKREKKAQAEAAQAEKEKPSKVEEVRRAMAQAQREKRAAAER  
RWASLNIOSSSSSTS"  
18283..19920  
/gene="F22M8.7"  
join(18283..18285,18409..18483,18880..19041,19459..19617,  
19837..19920)  
/gene="F22M8.7"  
/notes="Contains similarity to cyclophilin E from  
*Dicystosellum discoideum* gb|AF215865 and contains a  
Cyclophilin type peptidyl-prolyl cis-trans isomerase  
Pf|00160 domain."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAF76472.1"  
Query Match 78.3%; Score 18.8; DB 8; Length 47295;  
Best Local Similarity 68.2%; Pred. No. 1e+02;  
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 2 cauccuacuccaaucauucc 23  
II:II: I:IIII:IIII:II  
Db 14594 CATCCTTCTCCCATCCCAAGTCC 14573  
RESULT 9  
AP004529/c  
LOCUS 78249 bp DNA linear PLN 14-DEC-2001  
DEFINITION Lotus japonicus genomic DNA, chromosome 1, clone:LJT01K12, TM0058a,  
complete sequence.  
ACCESSION AP004529  
VERSION AP004529.1 GI:17736896  
KEYWORDS HTG.  
SOURCE Lotus japonicus DNA, clone:ljT library clone:ljT01K12.  
ORGANISM Lotus japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;  
Lotus.  
REFERENCE 1 (sites)  
AUTHORS Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.  
TITLE Structural Analysis of a Lotus japonicus Genome. I. Sequence  
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb  
Regions of the Genome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 78249)  
AUTHORS Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research, 1532-3, Yana,  
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,  
URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,





JOURNAL  
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15625694.

COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GCSI  
Center clone name: CH230-7021  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findhaplist  
Consensus quality: 112126 bases at least Q40  
Consensus quality: 120850 bases at least Q30  
Consensus quality: 128435 bases at least Q20  
Estimated insert size: 118743; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-tip estimation  
Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 4143: contig of 4143 bp in length  
\* 1  
\* 4144 4243: gap of unknown length  
\* 4244 9274: contig of 5031 bp in length  
\* 9275 9374: gap of unknown length  
\* 9375 15325: contig of 5951 bp in length  
\* 15326 15425: gap of unknown length  
\* 15426 22207: contig of 6782 bp in length  
\* 22207 22307: gap of unknown length  
\* 22308 25695: contig of 3388 bp in length  
\* 25695 25795: gap of unknown length  
\* 25796 31599: contig of 5804 bp in length  
\* 31600 31699: gap of unknown length  
\* 31700 35413: contig of 3714 bp in length  
\* 35414 35513: gap of unknown length  
\* 35514 39478: contig of 3965 bp in length  
\* 39479 39578: gap of unknown length  
\* 39579 43076: contig of 3498 bp in length  
\* 43077 43176: gap of unknown length  
\* 43177 45443: contig of 2267 bp in length  
\* 45444 45543: gap of unknown length  
\* 45544 48215: contig of 2672 bp in length  
\* 48216 48315: gap of unknown length  
\* 48316 51300: contig of 2985 bp in length  
\* 51301 51400: gap of unknown length  
\* 51401 53322: contig of 1922 bp in length  
\* 53323 53422: gap of unknown length  
\* 53423 56005: contig of 2583 bp in length  
\* 56006 56105: gap of unknown length  
\* 56106 58776: contig of 2671 bp in length  
\* 58777 58876: gap of unknown length  
\* 58877 60821: contig of 1945 bp in length  
\* 60822 60921: gap of unknown length  
\* 60922 62970: contig of 2049 bp in length  
\* 62971 63070: gap of unknown length  
\* 63071 64702: contig of 1632 bp in length  
\* 64703 64802: gap of unknown length  
\* 64803 67527: contig of 2725 bp in length  
\* 67528 71110: gap of unknown length  
\* 71110 71210: contig of 3483 bp in length  
\* 71211 71210: gap of unknown length  
\* 71211 74105: contig of 2895 bp in length

\* 74106 74205: gap of unknown length  
\* 74206 75907: contig of 1702 bp in length  
\* 75908 76007: gap of unknown length  
\* 76008 79435: contig of 3428 bp in length  
\* 79436 79535: gap of unknown length  
\* 79536 82171: contig of 2636 bp in length  
\* 82172 82272: gap of unknown length  
\* 82272 84362: contig of 2091 bp in length  
\* 84363 84462: gap of unknown length  
\* 84463 86374: contig of 1912 bp in length  
\* 86375 86474: gap of unknown length  
\* 86475 88823: contig of 2349 bp in length  
\* 88824 88923: gap of unknown length  
\* 88924 90852: contig of 1929 bp in length  
\* 90853 90952: gap of unknown length  
\* 90953 93358: contig of 2406 bp in length  
\* 93359 93458: gap of unknown length  
\* 93459 94986: contig of 1528 bp in length  
\* 94987 95086: gap of unknown length  
\* 95087 96893: contig of 1807 bp in length  
\* 96894 96993: gap of unknown length  
\* 96994 98840: contig of 1847 bp in length  
\* 98841 98940: gap of unknown length  
\* 98941 100977: contig of 2037 bp in length  
\* 100978 101077: gap of unknown length  
\* 101078 102920: contig of 1843 bp in length  
\* 102921 103020: gap of unknown length  
\* 103021 104528: contig of 1508 bp in length  
\* 104529 104628: gap of unknown length  
\* 104630 106895: contig of 2267 bp in length  
\* 106896 106995: gap of unknown length  
\* 106996 108325: contig of 1330 bp in length  
\* 108326 108425: gap of unknown length  
\* 108426 109446: contig of 1021 bp in length  
\* 109447 109546: gap of unknown length  
\* 109547 111148: contig of 1602 bp in length  
\* 111149 111248: gap of unknown length  
\* 111249 113082: contig of 1834 bp in length  
\* 113083 113182: gap of unknown length  
\* 113183 114745: contig of 1563 bp in length  
\* 114746 114845: gap of unknown length  
\* 114846 116015: contig of 1170 bp in length  
\* 116016 116115: gap of unknown length  
\* 116116 118467: contig of 2352 bp in length  
\* 118468 118567: gap of unknown length  
\* 118568 120547: contig of 1980 bp in length  
\* 120548 120647: gap of unknown length  
\* 120648 122231: contig of 1584 bp in length  
\* 122232 122331: gap of unknown length  
\* 122332 123520: contig of 1189 bp in length  
\* 123521 123620: gap of unknown length  
\* 123621 125501: contig of 1881 bp in length  
\* 125502 125601: gap of unknown length  
\* 125602 126912: contig of 1311 bp in length  
\* 126913 127012: gap of unknown length  
\* 127013 128420: contig of 1408 bp in length  
\* 128421 128520: gap of unknown length  
\* 128521 129699: contig of 1179 bp in length  
\* 129700 129799: gap of unknown length  
\* 129800 130901: contig of 1102 bp in length  
\* 130902 131001: gap of unknown length  
\* 131002 132646: contig of 1645 bp in length  
\* 132647 132746: gap of unknown length  
\* 132747 134251: contig of 1505 bp in length  
\* 134252 134351: gap of unknown length  
\* 134352 135499: contig of 1148 bp in length  
\* 135500 135599: gap of unknown length  
\* 135600 137135: contig of 1536 bp in length  
\* 137136 137235: gap of unknown length  
\* 137236 138470: contig of 1235 bp in length  
\* 138471 138470: gap of unknown length

Query Match

78.3%; Score 18.8; DB 2; Length 142898;

Best Local Similarity 63.6%; Pred. No. 1e+02; Matches 14; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 cauccuacucccauccaaaucc 23  
||:||||:||||:||||:  
Db 107853 CATCCTCTCCATCCAGTTC 107874

RESULT 12  
AC084272/c  
LOCUS AC084272 200768 bp DNA linear HTG 29-JAN-2002  
DEFINITION Mus musculus chromosome 16 clone rp23-11g21, WORKING DRAFT  
SEQUENCE, 8 unordered pieces.  
ACCESSION AC084272 GI:18390233  
VERSION AC084272.20 GI:18390233  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 200768)  
Do.T., Do.A. and Roe.B.A.  
Mus musculus Chromosome 16 BAC Clone rp23-11g21  
Unpublished  
2 (bases 1 to 200768)  
Do.T., Do.A. and Roe.B.A.  
Direct Submission  
Submitted (19-OCT-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Jan 29, 2002 this sequence version replaced gi:18139381.

## COMMENT

----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3097: contig of 3097 bp in length  
\* 3098 3197: gap of unknown length  
\* 3198 7898: contig of 4701 bp in length  
\* 7899 7898: gap of unknown length  
\* 7999 17599: contig of 9601 bp in length  
\* 17600 17699: gap of unknown length  
\* 17700 36839: contig of 19140 bp in length  
\* 36840 36939: gap of unknown length  
\* 36940 68422: contig of 31483 bp in length  
\* 68423 68523: gap of unknown length  
\* 68523 111403: contig of 42881 bp in length  
\* 111404 141504: gap of unknown length  
\* 141504 149778: contig of 38275 bp in length  
\* 149779 149878: gap of unknown length  
\* 149879 200768: contig of 50890 bp in length.

## FEATURES

## source

Location/Qualifiers  
1..200768  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="16"  
/clone="rp23-11g21"  
/clone\_lib="rp23"  
56087 a 46789 c 45810 g 51371 t 711 others

BASE COUNT  
ORIGIN

Query Match 78.3%; Score 18.8; DB 2; Length 200768;  
Best Local Similarity 63.6%; Pred. No. 1e+02;  
Matches 14; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 cauccuacucccauccaaaucc 23  
||:||||:||||:||||:  
Db 189911 CATCTAATCCCTCCAAATTC 189890

RESULT 13  
AC068162

LOCUS AC068162 203504 bp DNA linear HTG 09-JUN-2000  
DEFINITION Homo sapiens chromosome 3 clone RP11-543P4 map 3, WORKING DRAFT  
SEQUENCE, 31 unordered pieces.  
ACCESSION AC068162  
VERSION AC068162.2 GI:8389502  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 203504)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 3, clone RP11-543P4  
Unpublished  
2 (bases 1 to 203504)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Miengo,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J.,  
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J.,  
Yung,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 9, 2000 this sequence version replaced gi:7670156.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: t10146  
Center clone name: 543\_P\_4  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 184552 bases at least Q40  
Consensus quality: 194359 bases at least Q30  
Consensus quality: 198180 bases at least Q20  
Insert size: 208000; agarose-fp  
Insert size: 200504; sum-of-contigs  
Quality coverage: 3.8 in Q20 bases; agarose-fp  
Quality coverage: 4.0 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1      1303: contig of 1303 bp in length
*      1304 1403: gap of 100 bp
*      1404 2468: contig of 1065 bp in length
*      2469 2568: gap of 100 bp
*      2569 3769: contig of 1201 bp in length
*      3770 3869: gap of 100 bp
*      3870 5347: contig of 1478 bp in length
*      5348 5447: gap of 100 bp
*      5448 6492: contig of 1045 bp in length
*      6493 6592: gap of 100 bp
*      6593 8593: contig of 2001 bp in length
*      8594 8693: gap of 100 bp
*      8694 10995: contig of 2302 bp in length
*      10996 11095: gap of 100 bp
*      11096 13032: contig of 1937 bp in length
*      13033 13132: gap of 100 bp
*      13133 17251: contig of 4119 bp in length
*      17252 17351: gap of 100 bp
*      17352 20481: contig of 3130 bp in length
*      20482 20581: gap of 100 bp
*      20582 23646: contig of 3065 bp in length
*      23647 23746: gap of 100 bp
*      23747 27501: contig of 3755 bp in length
*      27502 27601: gap of 100 bp
*      27602 33325: contig of 5724 bp in length
*      33326 33425: gap of 100 bp
*      33426 37218: contig of 3793 bp in length
*      37219 37318: gap of 100 bp
*      37319 42731: contig of 5413 bp in length
*      42732 42831: gap of 100 bp
*      42832 46787: contig of 3956 bp in length
*      46788 46887: gap of 100 bp
*      46888 53272: contig of 6385 bp in length
*      53273 53372: gap of 100 bp
*      53373 61153: contig of 7781 bp in length
*      61154 61253: gap of 100 bp
*      61254 68270: contig of 7017 bp in length
*      68271 68370: gap of 100 bp
*      68371 77795: contig of 9425 bp in length
*      77796 77895: gap of 100 bp
*      77896 85280: contig of 7385 bp in length
*      85281 85380: gap of 100 bp
*      85381 94950: contig of 9570 bp in length
*      94951 95050: gap of 100 bp
*      95051 103176: contig of 8126 bp in length
*      103177 103276: gap of 100 bp
*      103277 110866: contig of 7590 bp in length
*      110867 110966: gap of 100 bp
*      110967 120234: contig of 9268 bp in length
*      120235 120334: gap of 100 bp
*      120335 129289: contig of 8955 bp in length
*      129290 129389: gap of 100 bp
*      129390 139564: contig of 10175 bp in length
*      139565 139664: gap of 100 bp
*      139665 150765: contig of 11101 bp in length
*      150766 150865: gap of 100 bp
*      150866 167813: contig of 16948 bp in length
*      167814 167913: gap of 100 bp
*      167914 185753: contig of 17840 bp in length
*      185754 185853: gap of 100 bp
*      185854 203504: contig of 17651 bp in length.
```

## FEATURES

source

```
1. 203504
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
```

```
/clone="RP11-543P4"
/clone_id="RP11 Human Male BAC"
1. 1303
/note="assembly_fragment"
1404. 2468
/note="assembly_fragment"
2569. 3769
/note="assembly_fragment"
3870. 5347
/note="assembly_fragment"
5448. 6492
/note="assembly_fragment
clone_end:SP6
vector_side:right"
6593. 8593
/note="assembly_fragment"
8694. 10995
/note="assembly_fragment"
11096. 13032
/note="assembly_fragment"
13133. 17251
/note="assembly_fragment"
17352. 20481
/note="assembly_fragment"
20582. 23646
/note="assembly_fragment
clone_end:T7
vector_side:left"
23747. 27501
/note="assembly_fragment"
27602. 33325
/note="assembly_fragment"
33426. 37218
/note="assembly_fragment"
37319. 42731
/note="assembly_fragment"
42832. 46787
/note="assembly_fragment"
46888. 53272
/note="assembly_fragment"
53373. 61153
/note="assembly_fragment"
61254. 68270
/note="assembly_fragment"
68371. 77795
/note="assembly_fragment"
77896. 85280
/note="assembly_fragment"
85381. 94950
/note="assembly_fragment"
95051. 103176
/note="assembly_fragment"
103277. 110866
/note="assembly_fragment"
110967. 120234
/note="assembly_fragment"
120335. 129289
/note="assembly_fragment"
129390. 139564
/note="assembly_fragment"
139665. 150765
/note="assembly_fragment"
```

Query Match 78.3% Score 18.8; DB 2; Length 203504;

Best Local Similarity 68.2%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;

Matches 15; Conservative 5; Mismatches

Qy 2 caucuaucuccaauucc 23

|||:|||||:||||:||||

Db 65837 CAACTACTCCATCCACTTCC 65858

RESULT 14

```

AC068980
LOCUS          204991 bp      DNA      linear      HTG 05-OCT-2001
DEFINITION    Homo sapiens chromosome 3 clone RP11-552A14, WORKING DRAFT
AC068980
SEQUENCE      8 unordered pieces.
AC068980
VERSION      GI:15741365
KEYWORDS     HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 204991)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaral-Tunke,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karasson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Ogih,M., Okwuonu,G., Oranuyn,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.I.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 204991)
Worley,K.C.
Direct Submission
Submitted (15-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 25, 2001 this sequence version replaced gi:11128156.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: HBGY
Center clone name: RP11-552A14
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 14% of reads
Chemistry: Dye-terminator Big Dye: 86% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 196487 bases at least Q40
Consensus quality: 203036 bases at least Q30
Consensus quality: 206264 bases at least Q20
Estimated insert size: 205706; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 52522: contig of 52522 bp in length
* 52523 52622: gap of unknown length
* 52623 91359: contig of 38737 bp in length
* 91360 91459: gap of unknown length
* 91460 125166: contig of 33707 bp in length
* 125167 125266: gap of unknown length
* 125267 152897: contig of 27631 bp in length
* 152898 152997: gap of unknown length
* 152998 173601: contig of 20604 bp in length
* 173602 173701: gap of unknown length
* 173702 188647: contig of 14946 bp in length
* 188648 188747: gap of unknown length
* 188748 202570: contig of 13823 bp in length
* 202571 204991: gap of unknown length
* 204991 204991: contig of 2321 bp in length.

FEATURES
Location/Qualifiers
source          1..204991
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="RP11-552A14"

BASE COUNT      60973 a 40617 c 40715 g 61970 t      716 others
ORIGIN

Query Match      78.3%; Score 18.8; DB 2; Length 204991;
Best Local Similarity 68.2%; Pred. No. 1e+02;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      2 cauccuacucccauucc 23
        |||||:||||:||||:||||
Db 45115 CAACCTACTCCATCCACATCC 45136

RESULT 15
AC024447
LOCUS          204993 bp      DNA      linear      HTG 11-APR-2000
DEFINITION    Homo sapiens chromosome 3 clone RP11-479J2 map 3, WORKING DRAFT
AC024447
SEQUENCE      26 unordered pieces.
AC024447
VERSION      GI:7534015
KEYWORDS     HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 204993)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,

```

TITLE  
JOURNAL  
COMMENT

Fenestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,  
Galan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, C., Hages, B., Heatford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Landers, T., Laroque, K., Lehoczy, J., Levine, R.,  
Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M.,  
McEwan, P., McGuirk, A., McKernan, K., McHeaters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olyar, T. M.,  
Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,  
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,  
Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and  
Zody, M.

## Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 11, 2000 this sequence version replaced gi:1108243.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L6997

Center clone name: 479\_J\_2

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 184405 bases at least Q40  
Consensus quality: 195348 bases at least Q40  
Consensus quality: 195362 bases at least Q20  
Insert size: 202493; sum-of-ctrls  
Quality coverage: 3.5 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 1060 1159: gap of 1059 bp in length  
\* 1160 3098: contig of 1939 bp in length  
\* 3099 3198: gap of 100 bp  
\* 3199 4151: contig of 953 bp in length  
\* 4152 4251: gap of 100 bp  
\* 4252 6077: contig of 1826 bp in length  
\* 6078 6177: gap of 100 bp  
\* 6178 10158: contig of 3981 bp in length  
\* 10159 10258: gap of 100 bp  
\* 10259 13440: contig of 3182 bp in length  
\* 13441 13540: gap of 100 bp  
\* 13541 16429: contig of 2889 bp in length  
\* 16430 16529: gap of 100 bp  
\* 16530 19606: contig of 3077 bp in length  
\* 19607 19706: gap of 100 bp  
\* 19707 24259: contig of 4553 bp in length  
\* 24260 24359: gap of 100 bp  
\* 24360 28314: contig of 3955 bp in length  
\* 28315 28414: gap of 100 bp  
\* 28415 32584: contig of 4170 bp in length  
\* 32585 32684: gap of 100 bp  
\* 32685 37793: contig of 5109 bp in length  
\* 37794 37893: gap of 100 bp

FEATURES  
source

\* 37894 44505: contig of 6612 bp in length  
\* 44506 44605: gap of 100 bp  
\* 44606 50170: contig of 5565 bp in length  
\* 50171 50270: gap of 100 bp  
\* 50271 57212: contig of 6942 bp in length  
\* 57213 57312: gap of 100 bp  
\* 57313 68003: contig of 10691 bp in length  
\* 68004 68103: gap of 100 bp  
\* 68104 77484: contig of 9381 bp in length  
\* 77485 77584: gap of 100 bp  
\* 77585 86863: contig of 9279 bp in length  
\* 86864 86963: gap of 100 bp  
\* 86964 96495: contig of 9532 bp in length  
\* 96496 96595: gap of 100 bp  
\* 96596 107929: contig of 11334 bp in length  
\* 107930 108029: gap of 100 bp  
\* 108030 116856: contig of 8827 bp in length  
\* 116857 116956: gap of 100 bp  
\* 116957 128225: contig of 11269 bp in length  
\* 128226 128325: gap of 100 bp  
\* 128326 142566: contig of 14241 bp in length  
\* 142567 142666: gap of 100 bp  
\* 142667 160140: contig of 17474 bp in length  
\* 160141 160240: gap of 100 bp  
\* 160241 179301: contig of 19061 bp in length  
\* 179302 179401: gap of 100 bp  
\* 179402 204993: contig of 25592 bp in length.  
1. 204993  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3"  
/clone="RP11-479J2"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 1059  
/note="assembly-fragment"  
1160. 3098  
/note="assembly-fragment"  
3199. 4151  
/note="assembly-fragment"  
clone\_end:77  
vector\_side:left  
4252. 6077  
/note="assembly-fragment"  
6178. 10158  
/note="assembly-fragment"  
10259. 13440  
/note="assembly-fragment"  
13541. 16429  
/note="assembly-fragment"  
16530. 19606  
/note="assembly-fragment"  
19707. 24259  
/note="assembly-fragment"  
24360. 28314  
/note="assembly-fragment"  
28415. 32584  
/note="assembly-fragment"  
32685. 37793  
/note="assembly-fragment"  
37894. 44505  
/note="assembly-fragment"  
44606. 50170  
/note="assembly-fragment"  
50271. 57212  
/note="assembly-fragment"  
57313. 68003  
/note="assembly-fragment"  
68104. 77484  
/note="assembly-fragment"  
77585. 86863  
/note="assembly-fragment"

```
misc_feature 86964..96495
              /note="assembly_fragment"
misc_feature 96596..107929
              /note="assembly_fragment"
misc_feature 108030..116856
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:left"
misc_feature 116957..128225
              /note="assembly_fragment"
misc_feature 128326..142566
              /note="assembly_fragment"
misc_feature 142667..160140
              /note="assembly_fragment"
misc_feature 160241..179301
              /note="assembly_fragment"
misc_feature 179402..204993
              /note="assembly_fragment"
BASE COUNT 62968 a 39652 c 40375 g 59489 t 2509 others
ORIGIN
```

```
Query Match 78.3%; Score 18.8; DB 2; Length 204993;
Best Local Similarity 68.2%; Pred.No.1e+02;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 cauccuacucccauccauucc 23
   |||:|||||:||||:|
Db 81605 CAACCTACTCCCATCCACTTCC 81626
```

Search completed: June 2, 2002, 18:57:16  
Job time: 11056 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 19:59:57 ; Search time 3664.46 Seconds  
(without alignments)  
2541.411 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_725  
Perfect score: 690  
Sequence: 1 atggcctctcttgccctcca.....acagcctgacagggtatgtg 690

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	610	88.4	822	10 BG386074	BG386074 602455248
2	599.6	86.9	728	10 BG325755	BG325755 602424466
3	585	84.8	680	10 BE304667	BE304667 601105783
4	579.4	84.0	978	10 BG164062	BG164062 602341087
5	568.8	82.4	945	10 BE513091	BE513091 601171545
6	547.6	79.4	2829	11 AK004990	AK004990 Mus muscu
7	503.6	73.0	727	10 B1101652	B1101652 602887470
8	502	72.8	724	10 B1100253	B1100253 602885716
9	492	71.3	643	10 BG328625	BG328625 602427889
10	484	70.1	777	10 B1102100	B1102100 602887847
11	473.4	68.6	758	10 B1102679	B1102679 602888236
12	460.8	66.8	706	10 B1148156	B1148156 602912333
13	460	66.7	494	10 BG385562	BG385562 602453858
14	456.8	66.2	846	10 BF789255	BF789255 602105156
15	451.8	65.5	750	9 AW475316	AW475316 un64d02.y
16	437.6	63.4	1114	10 BF789836	BF789836 602105395
17	434.6	63.0	673	9 AI746723	AI746723 u106f09.y

18	426.2	61.8	754	10 BG972599	BG972599 602841188
19	424	61.4	787	10 BF781962	BF781962 602106220
20	422.8	61.3	804	10 BF784189	BF784189 602108052
21	422.4	61.2	704	10 B1330754	B1330754 602982419
22	414.4	60.1	659	9 AI789490	AI789490 uk99h03.y
23	406.8	59.0	674	10 B1102491	B1102491 602889685
24	402.2	58.3	856	10 BF541534	BF541534 602087906
25	394.8	57.2	569	10 BE667438	BE667438 151799 MA
26	389.6	56.5	886	10 BF032123	BF032123 601559727
27	381	55.2	628	9 AI788398	AI788398 uk58c09.y
28	378.8	54.9	892	10 BF781303	BF781303 602106979
29	361.4	52.4	555	9 AV604013	AV604013 AV604013
30	358.2	51.9	560	9 AI953481	AI953481 wq29b02.x
31	354.4	51.4	593	9 AI789119	AI789119 uk51c01.y
32	353	51.2	1030	10 BF788818	BF788818 602110526
33	348.8	50.6	808	10 BG469035	BG469035 602510779
34	347	50.3	770	9 AI173400	AI173400 ue91f05.y
35	344	49.9	551	9 AV605969	AV605969 AV605969
36	328.6	47.6	616	9 AW475430	AW475430 un66c07.y
37	323.6	46.9	403	9 AI116661	AI116661 ue28a02.y
38	322.6	46.8	514	9 AI788419	AI788419 uk58e09.y
39	313.6	45.4	517	10 BE480970	BE480970 166244 BA
40	307.8	44.6	611	10 BF785226	BF785226 602108423
41	294.8	42.7	757	10 B1143586	B1143586 602907418
42	291.8	42.3	520	9 AI746647	AI746647 u101g04.y
43	291.6	42.3	816	10 B1144992	B1144992 602909649
44	288.6	41.8	376	9 AW610976	AW610976 un74h05.y
45	283.2	41.0	361	9 AI048112	AI048112 ud67a10.y

ALIGNMENTS

RESULT 1  
LOCUS BG386074  
DEFINITION 602455248F1 NTH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4583675 5', mRNA linear EST 12-MAR-2001  
ACCESSION BG386074  
VERSION BG386074.1 GI:13279520  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 822)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CMI307 row: p column: 12  
High quality sequence stop: 746.  
Location/Qualifiers  
1. .822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4583675"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in

BASE COUNT  
ORIGIN

161 a	224 c	217 g	200 t
-------	-------	-------	-------

```
QY 541 ctctgttttcttctcatcccaagaaatcgctccaactactacgatgacctaccaagcc 600
DB 569 CTCTGTCTTTTCTCTCTCATCCAGAGAAATCGCTCCAACTACTACGATG-CTACCAAGCC 627
QY 601 caacctctgcccacaaagagctctccaaggccttggtcaacctccccaaagtcaagagtga 660
DB 628 CAAC--TCTTGCACACAGAGCTCTCCRAGGC--TGGTCAACTCCCAAAAGTCAAGAGTGAG 683
QY 661 ttaattctacagcctgacaggggtatgtg 690
DB 684 -TCAATTCCTACAG-CTGACAGGGTATGTG 711

RESULT 3
BE304667 680 bp mRNA linear EST 13-JUL-2000
LOCUS 601105783F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2988309 5',
DEFINITION mRNA sequence.
ACCESSION BE304667
VERSION BE304667.1 GI:9176036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM77 row: f column: 22
High quality sequence stop: 27
High quality sequence stop: 680.
FEATURES
Location/Qualifiers
1..680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2988309"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 146 a 189 c 182 g 163 t
ORIGIN
source
Query Match 84.8%; Score 585; DB 10; Length 680;
Best Local Similarity 100.0%; Pred. No. 2.9e-151;
Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 gtcggtgccagcattgtgacagcagttggctctcccaaggccctctgagtggaatgtgcc 165
DB 38 GTCGGTGCCAGCATTTGACAGCAGATTGGCTTTCACCAAGGGCCCTCGGATGGAATGTGCC 97
QY 166 acacacagcagggcatcacccagtgacatctatagacccttctggcctgccccct 225
DB 98 ACACAGCAGCAGGCATCACCCAGTGAGATCTATAGCACCCCTTCTGGGCTGCGCCGCT 157
```

```
QY 226 gacatccaggctcccaaggccatgatgtagcatccagtgcaatctcctcctggcctgc 285
DB 158 GACATCAGGCTGCCAGGCCCATGATGGTGACATCCAGTCAATCTCTCTCCCTGGCCTGC 217
QY 286 attatctctgtgtggtgcatgagatgcacagttcttgcaggaatcccgagccaaagac 345
DB 218 ATTATCTCTGTGTGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCAGAGCCAAAGAC 277
QY 346 agagtggtggttagcaggtgagtccttttctcatcctctgagggcctcctgggtattcctcct 405
DB 278 AGAGTGGCGGTAGCAGGTGGAGTCTTTTTCATCTCTGGAGCCCTCTGGGATTCATTCTCT 337
QY 406 gttgcctggaattctcatgggatactcagggacttctactcaccacttggctgacagc 465
DB 338 GTTGGCCTGGAATCTTCATGGGATCCTACGGGACTTCTACTCACCACTGGTGGCTGACAGC 397
QY 466 atgaatttgagattgagagagctcttttacttgggcaattattctcctgttccctg 525
DB 398 ATGAAATTTGAGATTGGAGAGGCTCTTTACTTGGGCATTATTCTTCCCTTGTCTCCCTG 457
QY 526 atagctggaatcatcctctctgtcttcttcctgctcatccacagagaaatcgctccaaactactac 585
DB 458 ATAGCTGGAATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 517
QY 586 gatgcctacaaagcccaaacctcttgcacaagagagctctccaaaggcctggtcaacctccc 645
DB 518 GATGCTTACCAAGCCCAACCTCTTGGCACAAAGGAGCTCTCCAAGGCTGTGTTCAACCTCCC 577
QY 646 aaagtcaagagtgagttcaattctctacagcctgacaggggtatgtg 690
DB 578 AAAGTCAAGAGTGAGTTCAATTCCTACGCTGACAGGGTATGTG 622

RESULT 4
BG164062 978 bp mRNA linear EST 06-FEB-2001
LOCUS 602341087F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4449212 5',
DEFINITION mRNA sequence.
ACCESSION BG164062
VERSION BG164062.1 GI:12670765
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 978)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10232 row: m column: 21
High quality sequence stop: 689.
FEATURES
Location/Qualifiers
1..978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4449212"
/clone_lib="NIH_MGC_89"
/tissue_type="hypermphroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
```

BASE COUNT 220 a 279 c 276 g 203 t  
ORIGIN  
Query Match 84.0%; Score 579.4; DB 10; Length 978;  
Best Local Similarity 94.8%; Pred. No. 1.2e-149;  
Matches 655; Conservative 0; Mismatches 26; Indels 10; Gaps 5;  
QY 1 atggcctctcttgctcctcaactgtggtgctaacctctgaagcctcttgaggc 60  
DB 25 ATGGCTCTCTTGGCCCTCCAACTGTGGGCTACATCTTAGGCTTGGGGC 83  
QY 61 aacatgttgccatctgctgctcccaagctgaaacaaagtctatgtcggcgagcatt 120  
DB 84 ACATCTGTGCTCCATGCTGCTCCAGCTGGAACAAAGTTTATGTGGGTCACATTT 143  
QY 121 gtgacagagcttgctctctcacaaggcctctgtgatagtgacacacagcagc 180  
DB 144 GTGACAGAGCTTGGCTTCTCCAGGGGCTCTGATGGAATGTGCCACACAGCAGC 203  
QY 181 atcaccagtgatgacatctatagacacctctgggctgcgcgtgacatccagctg 240  
DB 204 ATCACCAGGTGTACATCTATACACACCTCTGTGGGCTGCCGCTGACATCAGGCTGCC 263  
QY 241 caggcgaatggtgacatccagctgacatctcctcctgctgcatatctctgtgtg 300  
DB 264 CAGGCCATGTGTGATCAGATCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 323  
QY 301 ggcatagatgacacagctctctcacaaggaaaccagcacaagagtgcgatgca 360  
DB 324 GGCATGAGATGACACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 383  
QY 361 ggtgagctcttctcactcttgagagcctcctggatctctctgtgcttgatctt 420  
DB 384 GGTGGAGCTCTTTTACCTTGGAGGCTCTGCGATTCATCTGCTGCTGCTGCTGCTG 443  
QY 421 catggatctacaggaactctactcaacacagctgtgctgacagcagaatttgagatt 480  
DB 444 CATGGATCTCTAGGGAGCTTCTACTCACACTGGTGGCTGACACATGAATTTGAGATT 503  
QY 481 ggaagagctc-ttacttggcatatctctcctgtctctcctgatagtgatact 539  
DB 504 GAGAGAGCTCTTACTTGGGCAATATTCTTCCCTGTCTCTGATAGCTGGAATCAT 563  
QY 540 cctctgcttctcctgctcctcacaagaaatcgctcaactactagatgcttccaa 599  
DB 564 CCTCTGCTTTCTCTGCTCAATCCAGAGAAATGCTCAACTACTAGATGCTTCAAA--- 620  
QY 600 ccaactcttgccacaaggagctctcacaaggctgtgacactctcccaagtcaagatga 659  
DB 621 --GCCAAGCTTGAAGAGAGCTCTCCAAAG--CTGGTCAACCTCCAAAGGTCAGAG- 675  
QY 660 gtcaattctcagacgctgacagatgtg 690  
DB 676 GTTCATTTCTTACAGGCGCACAGGATATGTG 706  
RESULT 5  
LOCUS BE513091 945 bp mRNA linear EST 07-AUG-2000  
DEFINITION 601171545F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:3545171 5',  
ACCESSION BE513091  
VERSION BE513091  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
REFERENCE 1 (bases 1 to 945)  
AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: L10M241 row: 1 column: 12  
High quality sequence start: 36  
High quality sequence stop: 782.  
FEATURES  
source  
1. 945  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3545171"  
/clone\_lib="NIH\_MGC.15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOMB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the Laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 194 a 262 c 263 g 226 t  
ORIGIN  
Query Match 82.4%; Score 568.8; DB 10; Length 945;  
Best Local Similarity 93.0%; Pred. No. 1e-146;  
Matches 640; Conservative 0; Mismatches 42; Indels 6; Gaps 4;  
QY 1 atggcctctcttgctcctcaactgtggtctacatcttgagcctcttgaggc 60  
DB 187 ATGGCTCTCTTGGCTCCAACTGTGGGCTACATCTAGGCTTGTGGGCTTTGGGC 246  
QY 61 aacatgttgccatctgctcccaagctgaaacaaagtctatgtggtgctgagcatt 120  
DB 247 ACATCTGTGCTCCATGCTGCTCCCACTGGAACAAAGTTTATCTGCTGCTGCTGCTG 306  
QY 121 gtgacagagcttgctctcacaaggcctctgtatgaaatgtgacacacagcagaagc 180  
DB 307 GTGACAGCAGTTGGCTTCTCCAAAGGCTCTGTGATGGAATGTGCCACACAGCAGCG 366  
QY 181 atcaccagtgatgacatctatagacacctcttggtcctgcgtgacatccagctgc 240  
DB 367 ATCACCAGGTGTGACATCTATAGCACCTTCTGGGCTCTGCCCTGACATCCAGGCTGCC 426  
QY 241 caggcgaatggtgacacagctgacatctcctcctgctgagatctctctgtggtg 300  
DB 427 CAGGCCATGATGTGACATCTGCAAGATCCCGAGCAAGAGACAGATGGGGATAGCA 486  
QY 301 ggcatagatgacacagctctctcacaaggaaaccagcacaagagtgcgatgca 360  
DB 487 GGCATGAGATGACAGCTTCTGCGCAGAGATCCCGAGCAAGAGACAGATGGGGATAGCA 546  
QY 361 ggtgagctcttctcactcttgagagcctcctggatctatctctgtgcttgaaatt 420  
DB 547 GGTGAGATC-TTTTCACTCTGGAGGCTCTGAGATCACTTCTGCTGCTGCTGCTG 605  
QY 421 catggatctcagagagctctactcaacacagtggtcgcagacagatgaattga--ga 478  
DB 606 CATGGATCTTACGGGACTTCTACTACACATGCTGCTGCTGCTGCTGCTGCTGCTG 665  
QY 479 ttgagagagctcttactctggcatattctctcctgtctcctcctgataagctgata 538  
DB 666 TGGCAGAGGCTCTTACTTGGGCTTATTTCTTCCCTGCTTCCCTGATAGCTGGATCA 725  
QY 539 -tctctgcttctcgtctacatccagagaaatcgctccaactactacagatgctacaa 597

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCGCAACTCGAGTGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'-GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCGCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

Location/Qualifiers  
 1. .2829  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGD:MG1:1904439"  
 /db\_xref="taxon:10090"  
 /clone="1300013G06"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 153..845  
 /gene="Cldn2"  
 153..845  
 /note="Cldn2"  
 /genes="claudin 2"  
 data source:MGD, source key:MG1:1276110, evidence:ISS putative"  
 /codon\_start=1  
 /protein\_id="BAB23725.1"  
 /db\_xref="GI:12836595"  
 /translation="MASLGVLVGYILGLIGLSTIAMLPNWRSTSSVVGASIVTAV GFSGLWMECATSTGTDCDITLGLPADTQAAQAMMTSSAAMSLIACITISVVG RCTVFQDSRAKDRVAVGVGFILGILGFTPVANLHGLIDFVSLPVPDSMKFEI GEALYIGITISLAVGVILCFSCSPQGNRTNYDGYOQAPLATRSSPSRQAQPAK SEFNSVSLTGYV"  
 polyA\_signal  
 2808..2813  
 /note="putative"  
 2829  
 /note="putative"  
 polyA\_site  
 718 a 707 c 674 g 730 t

BASE COUNT  
 ORIGIN  
 Query Match 79.4%; Score 547.6; DB 11; Length 2829;  
 Best Local Similarity 87.1%; Pred. No. 1.3e-140;  
 Matches 601; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
 Qy 1 atggcctctcttggcctcccaactgtgggtacaccttaggcctcttggggtttgggc 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 153 ATGGCCTCCCTTGGCTCCAACTGGTGGGTACATCCTAGGCCTTTGGGGCTGTAGGC 212  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 61 acaatggttgcattgtctgtcccccagctggaaacaagaattcttatgtcgtgcacatt 120  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 213 ACATCCATTGGCATGCTGCTTCCCAACTGGCAACGAGTTCCTATGTGTGGCAGCAT 272  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 121 gtgacagcagttgcttctccaaggcctctgcatggaatgtgccacacacacacagc 180  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 273 GTGACGGGGTGGCTTTTCCAAAGGGCCTCTGGATGAGTGTGGCACACACACACAGC 332  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 181 atcaccagttgtgacattctatagcacccttctggtcctgccctgacatccagctgccc 240  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCGCAACTCGAGTGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'-GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCGCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

Location/Qualifiers  
 1. .2829  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGD:MG1:1904439"  
 /db\_xref="taxon:10090"  
 /clone="1300013G06"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 153..845  
 /gene="Cldn2"  
 153..845  
 /note="Cldn2"  
 /genes="claudin 2"  
 data source:MGD, source key:MG1:1276110, evidence:ISS putative"  
 /codon\_start=1  
 /protein\_id="BAB23725.1"  
 /db\_xref="GI:12836595"  
 /translation="MASLGVLVGYILGLIGLSTIAMLPNWRSTSSVVGASIVTAV GFSGLWMECATSTGTDCDITLGLPADTQAAQAMMTSSAAMSLIACITISVVG RCTVFQDSRAKDRVAVGVGFILGILGFTPVANLHGLIDFVSLPVPDSMKFEI GEALYIGITISLAVGVILCFSCSPQGNRTNYDGYOQAPLATRSSPSRQAQPAK SEFNSVSLTGYV"  
 polyA\_signal  
 2808..2813  
 /note="putative"  
 2829  
 /note="putative"  
 polyA\_site  
 718 a 707 c 674 g 730 t

BASE COUNT  
 ORIGIN  
 Query Match 79.4%; Score 547.6; DB 11; Length 2829;  
 Best Local Similarity 87.1%; Pred. No. 1.3e-140;  
 Matches 601; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
 Qy 1 atggcctctcttggcctcccaactgtgggtacaccttaggcctcttggggtttgggc 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 153 ATGGCCTCCCTTGGCTCCAACTGGTGGGTACATCCTAGGCCTTTGGGGCTGTAGGC 212  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 61 acaatggttgcattgtctgtcccccagctggaaacaagaattcttatgtcgtgcacatt 120  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 213 ACATCCATTGGCATGCTGCTTCCCAACTGGCAACGAGTTCCTATGTGTGGCAGCAT 272  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 121 gtgacagcagttgcttctccaaggcctctgcatggaatgtgccacacacacacagc 180  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 273 GTGACGGGGTGGCTTTTCCAAAGGGCCTCTGGATGAGTGTGGCACACACACACAGC 332  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 181 atcaccagttgtgacattctatagcacccttctggtcctgccctgacatccagctgccc 240  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCGCAACTCGAGTGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'-GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCGCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

Location/Qualifiers  
 1. .2829  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGD:MG1:1904439"  
 /db\_xref="taxon:10090"  
 /clone="1300013G06"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 153..845  
 /gene="Cldn2"  
 153..845  
 /note="Cldn2"  
 /genes="claudin 2"  
 data source:MGD, source key:MG1:1276110, evidence:ISS putative"  
 /codon\_start=1  
 /protein\_id="BAB23725.1"  
 /db\_xref="GI:12836595"  
 /translation="MASLGVLVGYILGLIGLSTIAMLPNWRSTSSVVGASIVTAV GFSGLWMECATSTGTDCDITLGLPADTQAAQAMMTSSAAMSLIACITISVVG RCTVFQDSRAKDRVAVGVGFILGILGFTPVANLHGLIDFVSLPVPDSMKFEI GEALYIGITISLAVGVILCFSCSPQGNRTNYDGYOQAPLATRSSPSRQAQPAK SEFNSVSLTGYV"  
 polyA\_signal  
 2808..2813  
 /note="putative"  
 2829  
 /note="putative"  
 polyA\_site  
 718 a 707 c 674 g 730 t

BASE COUNT  
 ORIGIN  
 Query Match 79.4%; Score 547.6; DB 11; Length 2829;  
 Best Local Similarity 87.1%; Pred. No. 1.3e-140;  
 Matches 601; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
 Qy 1 atggcctctcttggcctcccaact

[illegible]

BASE COUNT	142 a	196 c	195 g	194 t	Technologies. Note: this is a NCI-CGAP Library.   "
ORIGIN					
Query Match	73.0%; Score 503.6; DB 10; Length 727;				
Best Local Similarity	86.8%; Pred: NC.1.le-128;				
Matches 554; Conservative	0; Mismatches 84; Indels 0; Gaps 0;				
OY	1	atggccctcttggcctccaactgtgtggctacatccatagagcctcttgaggcttttgagc	60		
DB	90	atggccctcttggcctccaactgtgtggctacatccatagagcctcttgaggcttttgagc	149		
OY	61	acactgtgttgcacatgctgtctcccaagcttgcgaagaaagcttctatgtcgtgtccagcalt	120		
DB	150	acatccattggcctgctgtctcccaagcttgcgaagaaagcttctatgtcgtgtccagcalt	209		
OY	121	gtgacagcagcttggctcttcccaaggctcttgatgtgaattgtgccacacagcagaagc	180		
DB	210	gtgacagcagcttggctcttcccaaggctcttgatgtgaattgtgccacacagcagaagc	269		
OY	181	atcaccacagctgtgacatctatagacacctcttcggccttcgcgcgtgacatccagctgc	240		
DB	270	atcaccacagctgtgacatctatagacacctcttcggccttcgcgcgtgacatccagctgc	329		
OY	241	caggccatgattgtgacatccagctgtgacatctctcccttgccttgcctgacatctctt	300		
DB	330	caggccatgattgtgacatccagctgtgacatctctcccttgccttgcctgacatctctt	389		
OY	301	ggcagatgagctgacagctcttcttgcgaagaatcccgagcccaagacagagtgagcga	360		
DB	390	ggcagatgagctgacagctcttcttgcgaagaatcccgagcccaagacagagtgagcga	449		
OY	361	ggtgagctcttcttcaacctcttggaagccctctggagatcatctcgttgcctggaatctt	420		
DB	450	ggtgagctcttcttcaacctcttggaagccctctggagatcatctcgttgcctggaatctt	509		
OY	421	catgagatctacagggaacttctactacacacactgtgtgcttgacagatgaattgagatt	480		
DB	510	catgagatctacagggaacttctactacacacactgtgtgcttgacagatgaattgagatt	569		
OY	481	ggaagagctcttacttcttgagcattatcttccctgcttctccctatagatgaaatcattc	540		
DB	570	ggaagagctcttacttcttgagcattatcttccctgcttctccctatagatgaaatcattc	629		
OY	541	cctctgcttcttgcctacatccagagaagaatgctcccaactactagatgcttccacaagcc	600		
DB	630	cctctgcttcttgcctacatccagagaagaatgctcccaactactagatgcttccacaagcc	689		
OY	601	caacctcttgacacaagagagcttctccaaggcttggta	638		
DB	690	caacctcttgacacaagagagcttctccaaggcttggta	727		
RESULT 8					
BI100253					
LOCUS	BI100253	724 bp	mRNA	linear	EST 26-JUN-2001
DEFINITION	602885716r1 NCI-CGAP_Kid14 Mus musculus cDNA clone IMAGE:5041370				
ACCESSION	BI100253				
VERSION	BI100253.1	GI:1451146			
KEYWORDS					
SOURCE					
ORGANISM	house mouse.				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: Jeffrey E. Green, M.D.				

```

|||||
Db 696 ATACAGCCTGACTGGGTATGTG 717

RESULT 9
BG328625 602427889f1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4547230 5',
LOCUS mRNA sequence. 643 bp mRNA linear EST 27-FEB-2001
DEFINITION
ACCESSION BG328625
VERSION BG328625.1 GI:13135063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 643)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LUCM1235 row: a column: 23
High quality sequence stop: 641.
FEATURES
source
1..643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4547230"
/clone_1lb="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 130 a 173 c 177 g 163 t
ORIGIN

Query Match 71.3%; Score 492; DB 10; Length 643;
Best Local Similarity 99.4%; Pred. No. 1.6e-125;
Matches 525; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 atggcctctcttgccctcccaacttggtggctacatcctagagcctcttggtggttttggtc 60
Db 116 ATGGCCCTCTCTGGCCCTCCCACTTGTGGGCTACATCCTAGGCCCTTCGTGGGCTTTTGGGC 175
QY 61 acactggttgccatgctgtctcccccagctggaacaaagaagttcttatgctggtgcagc-at 119
Db 176 ACACCTGGTGGCAGTGTCTCCAGGCGCTTGGAAAAACAAGTCTTATGTGCTGGTCGAGCGAT 235
QY 120 tgtacagcagcttggtgtcttcccaaggcctctgtaggaatgtgcaaca-cacagcacag 178
Db 236 TGTGACAGCAGTTGGCTTCTCCAGGCGCTCTGGAATGGAATGTGCCACAGCAGCACAG 295
QY 179 gcatcaccagctgtagcatctatagaccctctctggccctgcccgtgcacatccagctg 238
Db 296 GCATCACCAGTGTGACATCTATAGCACCCCTTCGTGGGCGCTGCCCGCTGCATCCAGGCTG 355
QY 239 cccaggccatgagtgtgacatccagtgcaatctctccctggcctgca-ttatctctgtg 297

```

```
Db 356 CCAGGCCATGATGGTGACATCCAGTCAATCTCCTCCCTGGCCTGCAGTTATCTCTGTG 415
Qy 298 gtggcatgagatgcacagctcttctgcaggaatcccgagcaaaagacagagtgcgcta 357
Db 416 GTGGGATGAGATGCACAGTCTTCTGTCAGGAATCCCGAGCAAGACAGAGTGGCGTA 475
Qy 358 gcagtgagagctcttttccatccctgagagcctcctggattcattcctgttgcctggaat 417
Db 476 GCAGGTGGAGTCTTTTCACTCTGGAGGCTCTCGGATTCATCTCTGTGGCTGGAAAT 535
Qy 418 ctctcaggtacatcagggactcttactcaccactggtgcctgacagcatgaaatttgag 477
Db 536 CTTCATGGGATCTACGGGACTTCTACTCACCAGTGGCTGACAGCATGAATTTGAG 595
Qy 478 attggagagcctcttacttggcattattcttccctgttccctcg 525
Db 596 ATTGGAGAGGCTCTTACTTGGGCATTTATTTCTTCCCTGTTCTCCCTG 643

RESULT 10
BI102100
LOCUS 602887847F1 NCI_CGAP_Kid14 777 bp mRNA linear EST 26-JUN-2001
DEFINITION 5', mRNA sequence.
ACCESSION BI102100
VERSION BI102100.1 GI:14552993
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1118 row: m column: 12
High quality sequence stop: 749.
FEATURES
Location/Qualifiers
1..777
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5043251"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 166 a 201 c 210 g 200 t
ORIGIN
```

Query Match 70.1%; Score 484; DB 10; Length 777;  
Best Local Similarity 86.3%; Pred. No. 2.9e-123;  
Matches 535; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy 1 atggcctcttggcctcccaactgtgggtacatcctagcctcttgggcttttgggc 60
Db 153 ATGGCCTCCCTGGCGTTCAACTGGTGGCTACATCCTAGGCTTTTGGGCTGTAGGC 212
Qy 61 acactggttccatgctgctcccgagctggaacacagttcttattgctgggcagcatt 120
Db 213 ACATCCATTGCCATGCTGCTTCCCACTGGCGAAGAGTTCCCTATGTGTGGTGCCAGCAT 272
```

```
Qy 121 gtgacagcagtggtgttcttcccaaggccctctgtggaatgtgccaacacacagcagc 180
Db 273 GTGACGGGGTGGTGGCTTTTCCAAAGGGCTCTGTGATGGAGTGTGGACACACAGCAGGC 332
Qy 181 atccccagtgagatctatagcacccttctggccttcccgctgacatccaggctgcc 240
Db 333 ATCCCCAGTGGGATATCTACAGTACCCCTTTAGGACTTCTGCTGACATCCAGGCTGCC 392
Qy 241 caggccatgagtgtgacatcccgagtgcaatctctccctggccctgcatattctctgtggtg 300
Db 393 CAGGCCATGATGATGACGTCCAGTGCAATGTCTCGCTGGCTTGTATTATCTCTGTGGTG 452
Qy 301 ggcagtgagatgcacagtgcttctgcagggaatcccgagccaaagacagagtggcggtagca 360
Db 453 GGCATGAGATGCACCGTGTCTGCCAGGATCTCGAGCTTAAGGACAGAGTGGCTGTAGTG 512
Qy 361 ggtgagtgcttttctcctctgagggccctcctgggattcattcctctgttgcctggaattct 420
Db 513 GGTGGAGTCTTTTTCATCTCTGCTGGCATCTCTGCGGCTTTATCCAGTTGCTTGGAAATCTT 572
Qy 421 catggatcctcagggactcttactcaccactggtgcctgacagcatgaaatttgagatt 480
Db 573 CATGGCATCTCTTCCGGGACTTCTACTCGCGCTGGTTCCTGACAGCATGAATTTGAGATT 632
Qy 481 ggagaggtcttcttacttggcattattcttccctgttctcctctgatatgctggaattcattc 540
Db 633 GGAGAGGCTCTCTACTTGGGCATCATCTCAGCCCTGTATACTATGTCGCGGAGTCAATC 692
Qy 541 ctctcgttctctgctcctcctcagagaaatgcctcacaactactacatcagctcctaccagcc 600
Db 693 CTTCCTTTTCTGCTGCTCGCCCGGCAATCGTACCACTACTATGATGCTGCTTAACAGGCC 752
Qy 601 caacctcttgccacagagag 620
Db 753 CAGCTTCTTGGCCACTAGGAG 772

RESULT 11
BI102679
LOCUS 602888236F1 NCI_CGAP_Kid14 758 bp mRNA linear EST 26-JUN-2001
DEFINITION 5', mRNA sequence.
ACCESSION BI102679
VERSION BI102679.1 GI:14553572
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1119 row: g column: 10
High quality sequence stop: 756.
FEATURES
Location/Qualifiers
1..758
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5043489"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)";
```

BASE COUNT  
ORIGIN

/note="Organ: kidney; Vector: PCMV-SPORT6; site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library. 1"

149 a 205 c 207 g 197 t

## Query Match

Best Local Similarity 66.8%; Score 473.4; DB 10; Length 756;  
Matches 547; Conservative 0; Mismatches 86; Indels 2; Gaps 2;

## Query Match

Best Local Similarity 86.1%; Pred. No. 2.5e-120;  
Matches 547; Conservative 0; Mismatches 86; Indels 2; Gaps 2;

COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs@emil.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L14M1145 row: a column: 09  
High quality sequence stop: 705.  
location/Qualifiers  
1..706  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5053328"  
/clone\_1ib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: PCMV-SPORT6; site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

FEATURES  
source

1 atggcctctcttgctcccaactgtgtggtacatccctagagctcttggtggtc 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
115 ATGGCCTCTCTTGCTCCCACTGTGAGGCTACATCCATAGGCTTGGGCTTGA 174  
61 acatgtgttcacatgctgtcctccagctggaagaacagtctatgtcgtgccaagcatt 120  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
175 ACATGTCATGGCATGCTCTCTCCACATGCGGACAGAGTCTCTATGTTGTGCCAGATT 234  
121 gtgaacagatgtgctctcccaaggcctctggaatgtgtgacacacagacagagc 180  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
235 GTGACGGCGGTGGCTTTCACAGGGCCTTGATGATGATGACACACAGACAGGC 294  
181 ataccacagatgtgaatctatagacacctcttggtcctgcccgtgacatcagctgcc 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
295 ATCACCCAGTCCGATATCTTACAGTACCCCTTTAGGATCCGACATCCAGGCTGCC 354  
241 caggccatgagtgagacatcagatccctccctgagcctgctgattctctgtgtg 300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
355 CAGGCCATGATGATGATCTCCACTGCAATGTCCTGCTGGCTTGTATATCTGTGGTG 414  
301 ggcataagatgcacagctctctcccaaggatcccgacccaagaagagtgagcgtagca 360  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
415 GGCAATGAGATGCACCGTCTTCCGACAGATCTGAGCTAAGACAGATGCTGTAGTG 474  
361 ggtgagatctcttcacatcccttggaagcctcctggaatcattctgtgtgcaatctt 420  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
475 GGTGAGAGCTTTTTCATCTGTGTCGACATCCGCTTTATCCAGTCTGTGAAATCTT 534  
421 catggagatcccaaggatctctacacacctgtgtcctgagaaagatgaattga-9at 479  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
535 CATGGCATCTTCCGGGATCTTACTGCGCGGTCTCGACAGCATGAATTTGACGAT 594  
480 tggagagcctcttactctgagcattatcttccctgtctccctgataagtgagatcc 539  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
595 TGGAGAGCTCTGTACTTGGGATCATCTCAGCCCTGTTTCTTGTGAGCGGAGTCA 654  
540 cctctgctcttctcagatcccaagaatcgtcccaactactaagatgctacaaagc 599  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
655 CCTTGTCTTCTTCTGCTGCGCCAGGTCATGCTACCAACTGATGATGCTACACAGC 714  
600 ccaacctcttgcacaagaagatctcccaagcctg 634  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
715 CC-AGCTCTTGCACCTAAGAGCTCTCACAAGATCTG 748

BASE COUNT 134 a 200 c 186 g 186 t  
ORIGIN

Query Match  
Best Local Similarity 66.8%; Score 460.8; DB 10; Length 706;  
Matches 574; Conservative 0; Mismatches 97; Indels 5; Gaps 5;

Query Match  
Best Local Similarity 84.9%; Pred. No. 7.5e-117;  
Matches 574; Conservative 0; Mismatches 97; Indels 5; Gaps 5;

1 atggcctctcttgctcccaactgtgtggtacatccctagagcctcttggtggtc 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
115 ATGGCCTCTCTTGCTCCCACTGTGAGGCTACATCCATAGGCTTGGGCTTGA 174  
61 acatgtgttcacatgctgtcctccagctggaagaacagtctatgtcgtgccaagcatt 120  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
175 ACATGTCATGGCATGCTCTCTCCACATGCGGACAGAGTCTCTATGTTGTGCCAGATT 234  
121 gtgaacagatgtgctctcccaaggcctctggaatgtgtgacacacagacagagc 180  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
235 GTGACGGCGGTGGCTTTCACAGGGCCTTGATGATGATGACACACAGACAGGC 294  
181 ataccacagatgtgaatctatagacacctcttggtcctgcccgtgacatcagctgcc 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
295 ATCACCCAGTCCGATATCTTACAGTACCCCTTTAGGATCCGACATCCAGGCTGCC 354  
241 caggccatgagtgagacatcagatccctccctgagcctgctgattctctgtgtg 300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
355 CAGGCCATGATGATGATCTCCACTGCAATGTCCTGCTGGCTTGTATATCTGTGGTG 414  
301 ggcataagatgcacagctctctcccaaggatcccgacccaagaagagtgagcgtagca 360  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
415 GGCAATGAGATGCACCGTCTTCCGACAGATCTGAGCTAAGACAGATGCTGTAGTG 474  
361 ggtgagatctcttcacatcccttggaagcctcctggaatcattctgtgtgcaatctt 420  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
475 GGTGAGAGCTTTTTCATCTGTGTCGACATCCGCTTTATCCAGTCTGTGAAATCTT 534  
421 catggagatcccaaggatctctacacacctgtgtcctgagaaagatgaattga-9at 479  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
535 CATGGCATCTTCCGGGATCTTACTGCGCGGTCTCGACAGCATGAATTTGACGAT 594  
480 tggagagcctcttactctgagcattatcttccctgtctccctgataagtgagatcc 539  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
595 TGGAGAGCTCTGTACTTGGGATCATCTCAGCCCTGTTTCTTGTGAGCGGAGTCA 654  
540 cctctgctcttctcagatcccaagaatcgtcccaactactaagatgctacaaagc 599  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
655 CCTTGTCTTCTTCTGCTGCGCCAGGTCATGCTACCAACTGATGATGCTACACAGC 714  
600 ccaacctcttgcacaagaagatctcccaagcctg 634  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
715 CC-AGCTCTTGCACCTAAGAGCTCTCACAAGATCTG 748

RESULT 12  
LOCUS B1148156 706 bp mRNA linear EST 05-JUL-2001  
DEFINITION 602912333F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5053328 5',  
mRNA sequence.  
ACCESSION B1148156  
VERSION B1148156.1 GI:14608157  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 706)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

1 atggcctctcttgctcccaactgtgtggtacatccctagagcctcttggtggtc 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
33 ATGGCCTCTCTTGCTCCCACTGTGAGGCTACATCCATAGGCTTGGGCTTGA 91  
61 acatgtgttcacatgctgtcctccagctggaagaacagtctatgtcgtgccaagcatt 120  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
92 ACATGTCATGGCATGCTCTCTCCACATGCGGACAGAGTCTCTATGTTGTGCCAGATT 149  
121 gtgaacagatgtgctctcccaaggcctctggaatgtgtgacacacagacagagc 180  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
150 GTGACGGCGGTGGCTTTCACAGGGCCTTGATGATGATGACACACAGACAGGC 209  
181 ataccacagatgtgaatctatagacacctcttggtcctgcccgtgacatcagctgcc 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
210 ATACCCAGTGCATATCTACAGTACCCCTTTAGAGATCTCTGTGACATCCAGCTGCC 269  
241 caggccatgagtgagacatcagatccctccctgagcctgctgattctctgtgtg 300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
270 CAGGCCATGATGATGATCTCCACTGCAATGTCCTGCTGGCTTGTATATCTGTGGTG 329  
301 ggcataagatgcacagctctctcccaaggatcccgacccaagaagagtgagcgtagca 360  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
330 GGCAATGAGATGCACCGTCTTCCGACAGATCTGAGCTAAGACAGATGCTGTAGTG 389  
361 ggtgagatctcttcacatcccttggaagcctcctggaatcattctgtgtgcaatctt 420  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
390 GGTGAGAGCTTTTTCATCTGTGTCGACATCCGCTTTATCCAGTCTGTGAAATCTT 449  
421 catggagatccctcagagatctcactcaacacagatgtgctgacagatgaattgagatt 480  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
450 CATGGCATCTTCCGGGATCTTACTGCGCGGTCTCGACAGCATGAATGATGAT 509  
481 ggaagagccttactctgagcattatcttccctgtctccct-9atagctggaatcatt 539  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
510 GGAGAGGCTCTGTACTGCGGACATCATCTCAGCCCTGCTCTCTGCTGCTGAGATCAT 569  
540 cctctgctcttctcagatcccaagaagaatcgtcccaactactaagatgctacaaagc 599  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
570 CCTTGTCTTCTCAGTCTGCCGCCAGGCAATGCTACCAACTGATGATGCTACAGGC 629  
600 ccaacctcttgcacaagaagatctcccaagcctgtgcaacctcccaagaatgaagattga 659  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db	630	CCAGCCTTTGCCACTAGGAGCTCTCCAAAGATCTGCTCAA-CAGCCAAAGCCAAAGATGA	688
Qy	660	gttcaattcctacagc	675
Db	689	GTTCAACTCATACAGC	704
RESULT 13			
LOCUS	602453859F1	NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4582214	5',
DEFINITION	494 bp	mRNA	linear EST 12-MAR-2001
ACCESSION	BG385562		mRNA sequence.
VERSION	BG385562.1	GI:13278350	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 494)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: NIH Intramural Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCMI304 row: c column: 15 High quality sequence stop: 493.		
FEATURES	Location/Qualifiers		
source	1..494		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4582214"		
	/clone_lib="NIH_MGC_15"		
	/tissue_type="adenocarcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"		
BASE COUNT	91 a 145 c 132 g 126 t		
ORIGIN			
Query Match	66.7%	Score 460;	DB 10;
Best Local Similarity	100.0%;	Pred. No. 1.le-116;	Length 494;
Matches	460;	Conservative 0;	Mismatches 0; Gaps 0;
Qy	1	atggcctcttggcctccactgtgggtctacatctctaggcctctgtgggcttttgggc	60
Db	35	ATGGCCTCTTGGCCTCCAACTTTGTGGGCTACATCCPAGGCCTTCTGGGCTTTTGGGC	94
Qy	61	acactgggtgcattgctgctcccgactggaaacaagtcttatgtcgtgccagcatt	120
Db	95	ACACTGGTTCATGCTGCTCCCAAGCTTGGATGGAAACAAGTCTTATGTGGTGCCAGCAT	154
Qy	121	gtgacagcagttggcttccaaaggcctctggatggaaatgtccacacacagcacaggc	180
Db	155	GTGACAGAGTTGGCTTCTCCAAAGGCCTCTGGATGGAAATGTGCCACACACAGCACAGC	214
Qy	181	atcaccagtgtagaatctatagcacccttctggcctgcccgcgtgacatccaggctgcc	240
Db	215	ATCACCAAGTGTGACATCTATAGACACCTCTTGGGCTTCGCGCTGACATCCAGCCTGCC	274



Search completed: June 2, 2002, 20:00:07  
Job time: 14617 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:39 ; Search time 108.28 Seconds  
(without alignments)  
1565.267 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_725  
Perfect score: 690  
Sequence: 1 atggcctctctggcctcca.....acagcctacaggatgtgt 690

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCRTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.8	10.7	363	2	US-08-966-316-8
2	70.8	10.3	413	3	US-09-188-930-55
3	43	6.2	7218	1	US-08-232-463-14
4	33.8	4.9	3468	1	US-07-951-715A-4
5	33.8	4.9	3468	2	US-07-951-715A-8
6	33.8	4.9	3468	2	US-08-459-448A-4
7	33.8	4.9	3468	2	US-08-459-448A-8
8	33.8	4.9	3468	3	US-08-459-595A-4
9	33.8	4.9	3468	3	US-08-459-595A-8
10	33.8	4.9	3468	3	US-08-459-504B-4
11	33.8	4.9	3468	3	US-08-459-504B-8
12	33.8	4.9	3468	3	US-08-459-444-4
13	33.8	4.9	3468	3	US-08-459-444-8
14	33.8	4.9	3468	3	US-09-053-549-5
15	33.8	4.9	3468	4	US-09-547-422-4
16	33.8	4.9	3468	4	US-09-547-422-8
17	33.8	4.9	3546	1	US-07-951-715A-10
18	33.8	4.9	3546	1	US-07-951-715A-12
19	33.8	4.9	3546	1	US-07-951-715A-14
20	33.8	4.9	3546	1	US-07-951-715A-27
21	33.8	4.9	3546	2	US-08-459-448A-10
22	33.8	4.9	3546	2	US-08-459-448A-12
23	33.8	4.9	3546	2	US-08-459-448A-14
24	33.8	4.9	3546	2	US-08-459-448A-27
25	33.8	4.9	3546	3	US-08-459-595A-10
26	33.8	4.9	3546	3	US-08-459-595A-12
27	33.8	4.9	3546	3	US-08-459-595A-14

28	33.8	4.9	3546	3	US-08-459-595A-27	Sequence 27, Appl
29	33.8	4.9	3546	3	US-08-459-504B-10	Sequence 10, Appl
30	33.8	4.9	3546	3	US-08-459-504B-12	Sequence 12, Appl
31	33.8	4.9	3546	3	US-08-459-504B-14	Sequence 14, Appl
32	33.8	4.9	3546	3	US-08-459-504B-27	Sequence 27, Appl
33	33.8	4.9	3546	3	US-08-459-444-10	Sequence 10, Appl
34	33.8	4.9	3546	3	US-08-459-444-12	Sequence 12, Appl
35	33.8	4.9	3546	3	US-08-459-444-14	Sequence 14, Appl
36	33.8	4.9	3546	3	US-08-459-444-27	Sequence 27, Appl
37	33.8	4.9	3546	4	US-09-547-422-10	Sequence 10, Appl
38	33.8	4.9	3546	4	US-09-547-422-12	Sequence 12, Appl
39	33.8	4.9	3546	4	US-09-547-422-14	Sequence 14, Appl
40	33.8	4.9	3546	4	US-09-547-422-27	Sequence 27, Appl
41	33.8	4.9	3547	1	US-07-951-715A-16	Sequence 16, Appl
42	33.8	4.9	3547	2	US-08-459-448A-16	Sequence 16, Appl
43	33.8	4.9	3547	3	US-08-459-595A-16	Sequence 16, Appl
44	33.8	4.9	3547	3	US-08-459-504B-16	Sequence 16, Appl
45	33.8	4.9	3547	3	US-08-459-444-16	Sequence 16, Appl

## ALIGNMENTS

RESULT 1  
US-08-966-316-8  
; Sequence 8, Application US/08966316  
; Patent No. 5932445  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Reddy, Roopa  
; APPLICANT: Murty, Lynn E.  
; APPLICANT: Mathur, Preeti  
; TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/966, 316  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36, 749  
; REFERENCE/DOCKET NUMBER: PF-0424 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 363 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: HNT3A2701  
; CLONE: 2417676  
; US-08-966-316-8

Query Match 10.7%; Score 73.8; DB 2; Length 363;  
Best Local Similarity 55.2%; Pred. No. 4e-14;  
Matches 144; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Qy 1 atggcctcttggcctccaaacttgggctacatcttaggccttctggtggttttgggc 60  
Db 48 ATGGCCTTCGCGGAATGAGATCCCTGGGAGTCTGCTGACACTGCTGGGTGGAAT 107

Qy 61 acactggttgcctgctgtctccagctggaacaaagtcttcttctgctggtgcccagcatt 120  
Db 108 GGCCTGGTCTCTGCCCCCTGCCCCATGTGGAAGTGACCGCTTTCATCGSCAACAGCATC 167

Qy 121 gtgacagcagttggtcttctccaggcctctggtggaatgtgcccacacacagcagc 180  
Db 168 GTGGTGGCCCAAGTGGTGTGGGAGGCGCTGTGGATGCTGCTGGTGTGACAGACCGCG 227

Qy 181 atcacccagtgtagactctatagcaccttctggtgctccgctgacatccaggtgcc 240  
Db 228 CAGATGCAGTGAAGGTGTACGACTACATGCTGGCGCTGCCACAGACCTGCAGGCTGCA 287

Qy 241 caggccatgattggtgacatcc 261  
Db 288 CGTGCCCTCTGTGTCATGCC 308

RESULT 2  
US-09-188-930-55  
; Sequence 55, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 55  
; LENGTH: 413  
; TYPE: DNA  
; ORGANISM: Human  
US-09-188-930-55

Query Match 10.3%; Score 70.8; DB 3; Length 413;  
Best Local Similarity 54.2%; Pred. No. 3.9e-13;  
Matches 166; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

Qy 199 tatagcaccttctggtgctgcccgtgacatccatcaggtgcccagccatgattggtgaca 258  
Db 21 tacgactgggtgctgcgcctgtctcggcctgtgagccactcagcctaatggtgtc 80

Qy 259 tcagtgcaatctctccctgctgctcattctctgtgtggtgcatgagatgcacagtc 318  
Db 81 tcctgtggtgctggtgcttctcgtgctgctgtgtgtgcccacgattggtgagtgcaagtcgcgc 140

Qy 319 ttctgcccaggaatcccgagccaaaga---cagagtggtgagcaggtgagtccttttc 375  
Db 141 tgtggggagacgacaagaatgaagaagccgctatagccatggtgagcagcataatttc 200

Qy 376 atccttggaggcctctggtgattcattctgttgcctggaattcttcattggtgatacgg 435  
Db 201 atcgtggcaggtcttgcgccttggtagctgtctctggtatggtgcatcagattgtcaca 260

Qy 436 gacttctactaccactggtgctgcatgacagcatgaaattgagattggagaggtctttac 495  
Db 261 gactttataaccctttgatccctaccacattaatgattgattgttggccctgccatctt 320

Qy 496 ttgggc 501  
Db 321 attggc 326

RESULT 3  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fis  
US-08-232-463-14

Query Match 6.2%; Score 43; DB 1; Length 7218;  
Best Local Similarity 6.0%; Pred. No. 0.0014;  
Matches 19; Conservative 169; Mismatches 129; Indels 0; Gaps 0;

Qy 329 aatcccgagcaaacagacagatggtggtgagtggtgagctcttctcctcttgaggcc 388  
Db 1029 AATCCGAGCTTGGCTGCGAGCTCGAGGAGCTGCGATVYVYVYVYVYVYVYVYVYVYV 1088

Qy 389 tcttgggattccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 448  
Db 1089 YV 1148

Qy 449 cactgggtgcctgacagcatgaaattgagattgagagagctcttacttgggcatattt 508  
Db 1149 YV 1208

Qy 509 ctccctgttctcctctgtagtggaatcctctctcttcttctcctcctcctccagagaa 568

DB 1209 YY 1268  
QY 569 atgcctcaactactacgtagctcaagaagcccaactcttgcacaaagagcttccaa 628  
DB 1269 YY 1328  
QY 629 ggcctgctcaactccc 645  
DB 1329 YYYYYYYYYYYYYYYY 1345

RESULT 4  
US-07-951-715A-4  
Sequence 4, Application US/07951715A  
Patent No. 5625136  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Roehstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttle, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/951,715A  
FILING DATE: 25-SEP-1992  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8615  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3468 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: 1, 3468  
OTHER INFORMATION: /product= "Full length synthetic  
maize optimized"  
OTHER INFORMATION: /note= "disclosed in Figure 3 as synfu1.mod. This sequenc  
OTHER INFORMATION: identical to flysnbl.fin as disclosed in Figure 1."  
US-07-951-715A-4

Query Match 4.9%; Score 33.8; DB 1; Length 3468;  
Best Local Similarity 48.7%; Pred. No. 0.83;  
Matches 92; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 59 gcaacactgttccatgctgtctcccaagctggaacaagttctatgtgtgtgcaaga 118  
DB 1403 GCAGCCAGATCACCCAGATCCCTGTGACCAAGACACCAACTGTGGCGCGGCACCGC 1462  
QY 119 ttgtgacagcagttgtcttccaaaggccttgatgagatltgcacacacagcaag 178  
DB 1463 TGTGTAGAGGCCCCCGCTTACCGGCGGACATCTGTGGCGCACACCGCCCGCCGAGA 1522  
QY 179 gcatcaccagtgatgacatctatagaccctctcgggctgcccgtcgtgacatccagctg 238  
DB 1523 TAGACACCTGGCGCGTGAACATCACCGCCCCCTGAGCCAGCGCTACCGGCTCGCATCC 1582  
QY 239 cccagcgca 247  
DB 1583 GCTACGCCA 1591

RESULT 5  
US-07-951-715A-8  
Sequence 8, Application US/07951715A  
Patent No. 5625136  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Roehstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttle, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/951,715A  
FILING DATE: 25-SEP-1992  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:

CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,448A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919)541-8582
: TELEFAX: (919)541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3468 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: HYPOTHETICAL: NO
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..3468
: OTHER INFORMATION: /product= "Full length synthetic
: OTHER INFORMATION: maize optimized"
: OTHER INFORMATION: /note= "Disclosed in Figure 3 as syntnl_mod. Th
: OTHER INFORMATION: identical to fisybnt.fin as disclosed in Figure
: US-08-459-448A-4
:

```

```

Query Match          4.9%;   Score 33.8;   DB 2;   Length 3468;
Best Local Similarity 48.7%;   Pred No. 0.83;
Matches 92; Conservative 0; Mismatches 97; Indels 0; Gaps

QY      59  gcacactggttgccatgtctctcccagctggaaaacaagtcttattcggtgccagca 118
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1403 GCAGCCAGATCACCCAGATCCCCCTGACCAAGCAGCACCAACTGGCGCAGCGCACGCG 1462

QY      119  ttgtgcacagcagttggtcttcctcaaggcgctctgatggaattgccccacacacagcag 178
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1463 TGTGTGAAGGGGCCCGCGCTTCACGGCGGGGAGACATCTTGC GGCGCACCAAGCCCGCGCCAGA 1522

QY      179  gcatcacccagtgtgacatctatagcacocctcttgggcgctgccgcgtgacatccaggctg 238
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1523 TCAGCACCTTGCGCGGTGAACATACCGCCCCCTTGAGCGAGCGGTACC GGTCCGCAATCC 1582

QY      239  cccaggcca 247
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1583 GCTACGCCA 1591

RESULT      7
US-08-459-448A-8
; Sequence 8, Application US/08459448A
; Patent No. 5859336
; GENERAL INFORMATION:
; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalin M.
```

RESULT 7  
US-08-459-448A-8  
; Sequence 8, Application US/08459448A  
; Patent No. 5859336  
; GENERAL INFORMATION:  
; APPLICANT: Kozziel, Michael G.  
; APPLICANT: Desai, Nalini M.







US-08-459-504B-8

Query Match	4.9%	Score 33.8;	DB 3;	Length 3468;
Best Local Similarity	48.7%;	Pred. No. 0.83;		
Matches	92;	Conservative	0;	Mismatches 97; Indels 0; Gaps 0;
QY	59	gcacactggttgccatctgctcccccagctggaacacaaagtctctatgtcggtgccacga	118	
Db	1403	GCAGCCAGATCACCCAGATCCCCCTGTACCAAGAGACCAACCTGGGCGAGCGGCCACGCG	1462	
QY	119	ttgtgacagcagttggtttctccaaggccctctctgtagtgnaatgtgcaacacacagcagacg	178	
Db	1463	TGSGTGAAGGGCCCCGGGTTTCACGGCGGCGGACATCTCTCGCGCGACCAAGCCCGGCCAGA	1522	
QY	179	gcataccccagttgacatctatagcacctctctggtccctgcgcgcgtgacatccaggtcg	238	
Db	1523	TCAGCACCCCTGGCGGTGAACATCACCGCCCCCTTGAGCGACGGCTACCGCGTCCGCATCC	1582	
QY	239	cccaggcca	247	
Db	1583	GCTACGCCA	1591	

RESULT 12  
US-08-459-444-4  
; Sequence 4, Application US/08459444A  
; Patent No. 6121014  
; GENERAL INFORMATION:  
; APPLICANT: Koziel, Michael G.  
; Desai, Nalini M.  
; Lewis, Kelly S.  
; Kramer, Vance C.  
; Warren, Gregory W.  
; Evola, Stephen V.  
; Crossland, Lyle D.  
; Wright, Martha S.  
; Merlin, Ellis J.  
; Launis, Karen L.  
; TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED  
; NUCLEIC ACID CODING SEQUENCE  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.

```

1  TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
2  NUCLEIC ACID CODING SEQUENCE
3
4  NUMBER OF SEQUENCES: 94
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: No. 6121014art1s Agribusiness Biotechnology
7  STREET: 3054 Cornwallis Road
8  CITY: Research Triangle Park
9  STATE: NC
10 COUNTRY: USA
11 ZIP: 27709
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent in Release #1.0, Version #1.30
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/459,444A
21 FILING DATE: 02-Jun-1995
22 CLASSIFICATION: <Unknown>
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 07/951,715
26 FILING DATE: 25-SEP-1992
27 APPLICATION NUMBER: US 07/772,027
28 FILING DATE: 04-OCT-1991
29
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Meligs, J. Timothy
32 REGISTRATION NUMBER: 38,241
33 REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV16
34
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (919)541-8587
37 TELEFAX: (919)541-8689
38
39 INFORMATION FOR SEQ ID NO: 4:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 3468 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..3468
; OTHER INFORMATION: /product= "Full length synthetic
; maize optimized"
; /note= "Disclosed in Figure 3 as synfl.mod. This sequence is
; identical to flsynbt.fin as disclosed in Figure 1."
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
;
; DS-08-459-444-4
;

```

Query Match	4.9%;	Score 33.8;	DB 3;	Length 3468;
Best Local Similarity	48.7%;	Pred. No. 0.83;		
Matches 92;	Conservative 0;	Mismatches 97;	Indels 0;	Gaps 0;
QY	59	gcacactggttgccatgctctcccaagctgagaaaacaagttcttatttcggtgccacga	118	
Db	1403	GCAGCCAGATCACCCAGATCCCTCTGACCAAGAGACCAACCTGGGGCAGCGCACACGG	1462	
QY	119	ttgtgcagcagttggtcttccaaaggcccttgatggaaatgtgtccacacacagcacag	178	
Db	1463	TGGTGAAGGGCCCGGCTTCCAGCGGGCGGACATCTCGCGCGCACCAAGCCCGGCCAGA	1522	
QY	179	gcatacccaagctgacatctatagcaccctctctggcctcccgctgacatcacaggctg	238	
Db	1523	TCAGCACCTCGCGGTGAACATCACCGCCCCCTCTGAGCGAGCGGTACCGCGGTCCGATCC	1582	
QY	239	cccaggcca	247	
Db	1583	GCTACGGCA	1591	

```

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
10
```

```

APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-18805/P1/GC1577/CIP/DIV6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3468 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3465
OTHER INFORMATION: /product= "Full-length, hybrid,
partially maize optimized cryIA(b)"
/note= "Disclosed in Figure 7 as contained in pCIB4434."
SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```

Query Match	4.9%	Score 33.8	DB 3	Length 3468
Best Local Similarity	48.7%	Pred. No. 0.83		
Matches	92	Conservative	0	Mismatches 97
				Indels 0
				Gaps 0
QY	59	gcacactggtgtccatgctcgtctctcccaagctgynaacaagttcttatgtcgtgcagca	118	
Db	1403	GCAGCCAGATATACCCAGATCCCCCTGACCACAGACCAACTGTGGCAGCGGCACACAGG	1462	
QY	119	tttfgacagcagtttggtctctccaaagggccctctgatatgaatgtgcacacacaaagcaag	178	
Db	1463	TGGTGAAGGGCCCCGGCTTCACCGGGCGGACATCCTGGCGCCGACACGCCCCGGCCAGA	1522	
QY	179	gcatacaccagttgtacatctataagcaacctctgtgagctgtccgctgacatcacagagctg	238	
Db	1523	TCAGACACCTCGCGGTGAATCACCGCCCCCTGAGCAAGCGCTACCGGCTCCGCATTC	1582	
QY	239	ccccagagcca	247	
Db	1583	GCTACGCCA	1591	

RESULT 14  
US-09-053-549-5  
Sequence 5, Application US/09053549  
Patent No. 6121521  
GENERAL INFORMATION:  
APPLICANT: Desai, Nalini  
TITLE OF INVENTION: No. 6121521a1 Insecticidal Protein and Gene  
NUMBER OF SPOUNCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6121521a1t1s Corporation  
STREET: 3054 Cornwallis Rd.  
City: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,549

```

1  FILING DATE: 01-APR-1998
2  CLASSIFICATION: 800
3  ATTORNEY/AGENT INFORMATION:
4      NAME: Pace, Gary M.
5      REGISTRATION NUMBER: 40,403
6      REFERENCE/DOCKET NUMBER: CCC 1995
7      TELECOMMUNICATION INFORMATION:
8      TELEPHONE: 919-541-8582
9      TELEFAX: 919-541-8689
10 INFORMATION FOR SEQ ID NO: 5:
11     SEQUENCE CHARACTERISTICS:
12         LENGTH: 3468 base pairs
13         TYPE: nucleic acid
14         STRANDEDNESS: single
15         TOPOLOGY: linear
16     MOLECULE TYPE: other nucleic acid
17     DESCRIPTION: /desc = "Synthetic DNA"
18     HYPOTHETICAL: NO
19     FEATURE:
20         NAME/KEY: misc_feature
21         LOCATION: 1..3468
22     OTHER INFORMATION: /product= "Full length synthetic
23     OTHER INFORMATION: maize optimized"
24     FEATURE:
25         NAME/KEY: CDS
26         LOCATION: 1..3468
27     ;
28 ;
29 US-09-053-549-5

```

	Query Match	Score	DB	Length
	4.9%	33.8	3	3468
	Best Local Similarity	48.7%	Pred. No.	0.83
	Matches	92	Conservative	0
			Mismatches	97
			Indels	0
			Gaps	0
Qy	59	gcaacatggtgtgcatagtctcctcccaagtgtgaaacaagtcttattgttcggtccagca	118	
Db	1403	GCAGCCAAATCACCACAGATCCCTTGACCAAGACACCAACTGGGGACAGGCACACGCG	1462	
Qy	119	ttgtgacagcagttgtgctcttccaaagtgctctgtatgtgaatgtgcacacacagcaagc	178	
Db	1463	TGgtTAAGGGCCCCGGCTTACCGGGGGGAGATCCTGTGGCCGCAACCAACCCCGGCCAGA	1522	
Qy	179	gcatcaaccagatgtgtacatctatagcacccttctgtgacctgcgctgtgacatccagctgt	238	
Db	1523	TCAGACACCTGGGGCTGAAATCACCACCGCCCCCTGTAGCCAGCGCTAACCGCTCCGATCC	1582	
Qy	239	cccagagcca	247	
Db	1583	GCTACGCCA	1591	

RESULT 15  
US-09-547-422-4  
Sequence 4, Application US/09547422  
Patent No. 6320100  
GENERAL INFORMATION:  
APPLICANT: Kozziel, Michael G.  
Desai, Nalini M.  
Lewis, Kelly S.  
Kramer, Vance C.  
Warren, Gregory W.  
Evola, Stephen V.  
Crossland, Lyle D.  
Wright, Martha S.  
Merlin, Ellis J.  
Launis, Karen L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6320100artis Agribusiness Biotechnology Research, Inc.  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 20:08:48 ; Search time 523.16 Seconds

(without alignments)  
2264.452 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_725

Perfect score: 690

Sequence: 1 atggcctctcttgctcca.....acagcctgacaggtatgtg 690

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	690	100.0	1400	20	AA97865
2	690	100.0	1475	21	AA37060
3	690	100.0	1475	22	AA515360
4	690	100.0	1475	22	AA521489
5	690	100.0	1475	22	AA592097
6	690	100.0	1475	22	AA54296
7	690	100.0	1481	22	AA508497
8	690	100.0	1918	24	AA04424
9	690	100.0	2742	21	AA12585

10	690	100.0	2863	22	AA93769	Human CDNA encoding
11	689.6	99.9	1441	24	AA516180	Human CASB81 polyn
12	689.2	99.9	1524	20	AA97818	Human secreted pro
13	688.4	99.8	1540	22	AA158013	Human polynucleoti
14	688.4	99.8	1761	22	AA159799	Human polynucleoti
15	546	79.1	779	24	AA516181	Murine CASB81 poly
16	546	79.1	791	21	AA289137	Murine clodrin 2 cd
17	417.4	60.5	530	22	AA088518	Human secreted pro
18	388.8	56.3	467	21	AA98817	Human pancreatic c
19	361.6	52.4	615	22	AA93963	Primer specific fo
20	339.2	49.2	405	20	AA41059	Human secreted pro
21	263	38.1	265	22	AA34986	Human colon cancer
22	251.8	36.5	1046	22	AA088519	Human secreted pro
23	215.8	31.3	1174	21	AA37113	Human PRO1571 (ONQ
24	215.8	31.3	1174	22	AA546101	Human DNA encoding
25	215.8	31.3	1174	22	AA54431	Probe #47 used in
26	196.6	28.5	1361	22	AAK51692	Human polynucleoti
27	193.4	28.0	1381	22	AAK52676	Human polynucleoti
28	165	23.9	182	22	AA67470	Novel human polynu
29	165	23.9	324	20	AA41060	Human secreted pro
30	154.8	22.4	1779	20	AAV80582	Kidney injury asso
31	146.8	21.3	766	23	AA592167	DNA encoding novel
32	143.8	20.8	300	20	AAZ13117	Human gene express
33	141	20.4	681	22	AA60261	Human Claudin-7 CD
34	141	20.4	1014	21	AA695568	Human secreted pro
35	141	20.4	1253	20	AA242051	Human secreted pro
36	141	20.4	1546	21	AA247929	Human endometrium
37	141	20.4	1583	22	AAH34904	Human apoptosis as
38	141	20.4	1711	19	AAV43617	Human colon cancer
39	141	20.4	1787	22	AA98719	Human secreted pro
40	139.4	20.2	1705	20	AAV84506	Human late stage o
41	139.4	20.2	1705	22	AA883289	Human secreted pro
42	138.8	20.1	1665	21	AA236702	Human secreted pro
43	138.8	20.1	1665	22	ABA83103	Nucleotide sequenc
44	138.8	20.1	1717	22	AA98710	Claudin 4 ovarian
45	138.8	20.1	1776	22	AAH34872	Human late stage o
						Human colon cancer

#### ALIGNMENTS

RESULT 1	
AA97865	standard; CDNA: 1400 BP.
ID	AA97865
AC	AA97865;
XX	
DT	23-SEP-1999 (first entry)
XX	
DE	Human secreted protein encoding CDNA #53.
XX	
KW	Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;
XX	diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
OS	Homo sapiens.
XX	
PN	MO9925825-A2.
XX	
PD	27-MAY-1999.
XX	
PF	13-NOV-1998;
XX	
PR	04-SEP-1998;
XX	
PR	13-NOV-1997;
XX	
PR	17-DEC-1997;
XX	
PR	09-FEB-1998;
XX	
PR	13-APR-1998;
XX	
XX	10-AUG-1998;
PA	(GENSET) GENSET.
XX	
PI	Bougueleret L, Duclet A, Dumas Milne Edwards J;
XX	

DR WPI: 1999-347472/29.  
DR P-PSDB; AAY36181.  
XX  
XX  
PT Extended cDNAs encoding secreted proteins  
XX  
XX  
XX Claim 1; Page 254-255; 307pp; English.  
XX  
XX AAX97813-X97906 represent extended cDNA's which encode novel human  
CC secreted proteins (see AAY36129-Y36222) and which have cytostatic,  
CC thrombotic and osteopathic activity. The extended cDNAs can be used to  
CC express secreted proteins or parts of them or to obtain antibodies  
CC capable of binding to the secreted proteins. They may also be used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC Uses also include design of expression vectors and secretion vectors.  
XX  
XX Sequence 1400 BP; 326 A; 392 C; 346 G; 336 T; 0 other;

Query Match 100.0%; Score 690; DB 20; Length 1400;  
Best Local Similarity 100.0%; Pred. No. 7.4e-204;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctctcttggcctcccaactgttggtggtacatccttagcctctctggggttttgggc 60  
Db |||||  
QY 36 atggcctctcttggcctcccaactgttggtggtacatccttagcctctctggggttttgggc 95  
Db |||||  
QY 61 acactgttgcctatgtctgtctccagctggaacaaagtcttattgtcgtgcccagcatt 120  
Db |||||  
QY 96 acactgttgcctatgtctgtctccagctggaacaaagtcttattgtcgtgcccagcatt 155  
Db |||||  
QY 121 gtgacagcagtggtctctccagggcctctggtggaatgtgccacacagcacagggc 180  
Db |||||  
QY 156 gtgacagcagtggtctctccagggcctctggtggaatgtgccacacagcacagggc 215  
Db |||||  
QY 181 atcaccagtggtgacatctatagaccctcttggtggtggtggtggtggtggtggtggtg 240  
Db |||||  
QY 216 atcaccagtggtgacatctatagaccctcttggtggtggtggtggtggtggtggtggtg 275  
Db |||||  
QY 241 caggccatgatgtgacatccagtcgaatctctccctggtggtggtggtggtggtggtggtg 300  
Db |||||  
QY 276 caggccatgatgtgacatccagtcgaatctctccctggtggtggtggtggtggtggtggtg 335  
Db |||||  
QY 301 ggcattgagatgacagctctctccaggaatcccgagccaaagacagagtggtggtgagca 360  
Db |||||  
QY 336 ggcattgagatgacagctctctccaggaatcccgagccaaagacagagtggtggtgagca 395  
Db |||||  
QY 361 ggtgagctcttttctccttgagcctctggtggtggtggtggtggtggtggtggtggtggt 420  
Db |||||  
QY 396 ggtgagctcttttctccttgagcctctggtggtggtggtggtggtggtggtggtggtggt 455  
Db |||||  
QY 421 catggatcctacgggacttctcaccactggtggtggtggtggtggtggtggtggtggtggt 480  
Db |||||  
QY 456 catggatcctacgggacttctcaccactggtggtggtggtggtggtggtggtggtggtggt 515  
Db |||||  
QY 481 ggagagctcttacttggcattatttctcctggtggtggtggtggtggtggtggtggtggtggt 540  
Db |||||  
QY 516 ggagagctcttacttggcattatttctcctggtggtggtggtggtggtggtggtggtggtggt 575  
Db |||||  
QY 541 ctctgcttttctcctgctcctccagagaaatcgctcctcactacgactgctcctcctcctcct 600  
Db |||||  
QY 576 ctctgcttttctcctgctcctccagagaaatcgctcctcactacgactgctcctcctcctcct 635  
Db |||||  
QY 601 caactcttgcacagggagctctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 660  
Db |||||  
QY 636 caactcttgcacagggagctctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 695  
Db |||||  
QY 661 ttcactctcctacgctgacagggatgtg 690  
Db |||||  
QY 696 ttcactctcctacgctgacagggatgtg 725  
Db |||||

AA37060 standard; cDNA; 1475 BP.  
AAA37060;  
08-AUG-2000 (first entry)  
Human PRO1356 (UNQ705) cDNA sequence SEQ ID NO:133.  
Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
transmembrane; secretion; immunoadhesion; pharmaceutical; screening;  
ss.  
Homo sapiens.  
WO200012708-A2.  
09-MAR-2000.  
01-SEP-1999; 99WO-US20111.  
01-SEP-1998; 98US-0098716.  
01-SEP-1998; 98US-0098749.  
01-SEP-1998; 98US-0098750.  
02-SEP-1998; 98US-0098803.  
02-SEP-1998; 98US-0098821.  
02-SEP-1998; 98US-0098843.  
02-SEP-1998; 98US-0099536.  
09-SEP-1998; 98US-0099596.  
09-SEP-1998; 98US-0099598.  
09-SEP-1998; 98US-0099602.  
09-SEP-1998; 98US-0099642.  
10-SEP-1998; 98US-0099741.  
10-SEP-1998; 98US-0099754.  
10-SEP-1998; 98US-0099763.  
10-SEP-1998; 98US-0099792.  
10-SEP-1998; 98US-0099808.  
10-SEP-1998; 98US-0099812.  
10-SEP-1998; 98US-0099815.  
10-SEP-1998; 98US-0099816.  
15-SEP-1998; 98US-0100385.  
15-SEP-1998; 98US-0100388.  
15-SEP-1998; 98US-0100390.  
16-SEP-1998; 98US-0100584.  
16-SEP-1998; 98US-0100627.  
16-SEP-1998; 98US-0100661.  
16-SEP-1998; 98US-0100662.  
16-SEP-1998; 98US-0100664.  
17-SEP-1998; 98US-0100683.  
17-SEP-1998; 98US-0100684.  
17-SEP-1998; 98US-0100710.  
17-SEP-1998; 98US-0100711.  
17-SEP-1998; 98US-0100919.  
17-SEP-1998; 98US-0100930.  
18-SEP-1998; 98US-0100848.  
18-SEP-1998; 98US-0100849.  
18-SEP-1998; 98US-0101014.  
18-SEP-1998; 98US-0101068.  
18-SEP-1998; 98US-0101071.  
22-SEP-1998; 98US-0101279.  
23-SEP-1998; 98US-0101471.  
23-SEP-1998; 98US-0101472.  
23-SEP-1998; 98US-0101474.  
23-SEP-1998; 98US-0101475.  
23-SEP-1998; 98US-0101476.  
23-SEP-1998; 98US-0101477.  
23-SEP-1998; 98US-0101479.  
24-SEP-1998; 98US-0101738.  
24-SEP-1998; 98US-0101741.  
24-SEP-1998; 98US-0101743.  
24-SEP-1998; 98US-0101915.  
24-SEP-1998; 98US-0101916.  
29-SEP-1998; 98US-0102207.  
29-SEP-1998; 98US-0102240.

PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102330.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.  
PR 30-SEP-1998; 98US-0102570.  
PR 30-SEP-1998; 98US-0102571.  
PR 01-OCT-1998; 98US-0102684.  
PR 01-OCT-1998; 98US-0102687.  
PR 01-OCT-1998; 98US-0102687.  
PR 02-OCT-1998; 98US-0102655.  
PR 06-OCT-1998; 98US-0103258.  
PR 06-OCT-1998; 98US-0103449.  
PR 07-OCT-1998; 98US-0103314.  
PR 07-OCT-1998; 98US-0103315.  
PR 07-OCT-1998; 98US-0103328.  
PR 07-OCT-1998; 98US-0103395.  
PR 07-OCT-1998; 98US-0103396.  
PR 07-OCT-1998; 98US-0103401.  
PR 08-OCT-1998; 98US-0103633.  
PR 08-OCT-1998; 98US-0103678.  
PR 08-OCT-1998; 98US-0103679.  
PR 08-OCT-1998; 98US-0103711.  
PR 14-OCT-1998; 98US-0104257.  
PR 20-OCT-1998; 98US-0104987.  
PR 20-OCT-1998; 98US-0105000.  
PR 20-OCT-1998; 98US-0105002.  
PR 21-OCT-1998; 98US-0105104.  
PR 22-OCT-1998; 98US-0105169.  
PR 22-OCT-1998; 98US-0105266.  
PR 26-OCT-1998; 98US-0105693.  
PR 26-OCT-1998; 98US-0105694.  
PR 26-OCT-1998; 98US-0105807.  
PR 27-OCT-1998; 98US-0105881.  
PR 27-OCT-1998; 98US-0105882.  
PR 27-OCT-1998; 98US-0106032.  
PR 28-OCT-1998; 98US-0106033.  
PR 28-OCT-1998; 98US-0106178.  
PR 28-OCT-1998; 98US-0106248.  
PR 29-OCT-1998; 98US-0106384.  
PR 29-OCT-1998; 98US-0106384.  
PR 29-OCT-1998; 98US-0106500.  
PR 30-OCT-1998; 98US-0106464.  
PR 03-NOV-1998; 98US-0106856.  
PR 03-NOV-1998; 98US-0106902.  
PR 03-NOV-1998; 98US-0106905.  
PR 03-NOV-1998; 98US-0106919.  
PR 03-NOV-1998; 98US-0106932.  
PR 03-NOV-1998; 98US-0106934.  
PR 10-NOV-1998; 98US-0107783.  
PR 17-NOV-1998; 98US-0108775.  
PR 17-NOV-1998; 98US-0108779.  
PR 17-NOV-1998; 98US-0108787.  
PR 17-NOV-1998; 98US-0108788.  
PR 17-NOV-1998; 98US-0108801.  
PR 17-NOV-1998; 98US-0108802.  
PR 17-NOV-1998; 98US-0108806.  
PR 17-NOV-1998; 98US-0108807.  
PR 17-NOV-1998; 98US-0108867.  
PR 17-NOV-1998; 98US-0108925.  
PR 18-NOV-1998; 98US-0108848.  
PR 18-NOV-1998; 98US-0108849.  
PR 18-NOV-1998; 98US-0108850.  
PR 18-NOV-1998; 98US-0108851.  
PR 18-NOV-1998; 98US-0108852.  
PR 18-NOV-1998; 98US-0108858.  
PR 18-NOV-1998; 98US-0108904.  
XX  
XX  
(GETH ) GENENTECH INC.  
PA Baker K, Goddard A, Gurney AL, Smith V, Matanabe CK, Wood WI;  
XX

XX WPI: 2000-237871/20.  
DR P-PSDB; AAY93378.  
XX  
XX New mammalian DNA sequences encoding transmembrane, receptor or  
PT secreted PRO polypeptides, useful for screening of potential peptide or  
PT small molecule inhibitors of the relevant receptor/ligand interactions  
XX  
XX Claim 2; Fig 77; 773pp; English.  
XX  
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,  
CC receptor or secreted PRO polypeptides given in AAY9340 to AAY9462. The  
CC transmembrane and receptor PRO proteins can be used for screening of  
CC potential peptide or small molecule inhibitors of the relevant  
CC receptor/ligand interactions. The polypeptides and nucleotide sequences  
CC encoding them have various industrial applications, including uses as  
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent  
CC PCR primers and hybridisation probes used in the isolation of the PRO  
XX polypeptides from the present invention.  
XX  
SQ Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;  
  
Query Match 100.0%; Score 690; DB 21; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.6e-204;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 atggcctctctggcctcaacttgggcttatcccttggccttgggcttggg 60  
DB 122 atggcctctctggcctcaacttgggcttatcccttggccttgggcttggg 181  
  
QY 61 acactgtgtccatgtctgtccccaacttggaaacagttctatgtcgtgcacatt 120  
DB 182 acactgtgtccatgtctgtccccaacttggaaacagttctatgtcgtgcacatt 241  
  
QY 121 gtgaacagcagttgtgcttccaaaggcctctgtgaatgtgtccacacacagcagc 180  
DB 242 gtgaacagcagttgtgcttccaaaggcctctgtgaatgtgtccacacacagcagc 301  
  
QY 181 atcaccacatgtgacatcatatagacaccttctgggcttcccgctatccacagctgtcc 240  
DB 302 atcaccacatgtgacatcatatagacaccttctgggcttcccgctatccacagctgtcc 361  
  
QY 241 caggcccatatgtgtgacacacagttgcaatctctccctggcctgcgcatatctgtgggtc 300  
DB 362 caggcccatatgtgtgacacacagttgcaatctctccctggcctgcgcatatctgtgggtc 421  
  
QY 301 ggcattgagatgacagctctctgtccaggaatcccgagccaagaagatgtgcgttagca 360  
DB 422 ggcattgagatgacagctctctgtccaggaatcccgagccaagaagatgtgcgttagca 481  
  
QY 361 ggttgagctctttcaaccttggaggcctcctgggattcaattcctgttgcgtgaactct 420  
DB 482 ggttgagctctttcaaccttggaggcctcctgggattcaattcctgttgcgtgaactct 541  
  
QY 421 catggatcctacgggacttctacacacatgtgtcccgacagcagtcgtaatttgatt 480  
DB 542 catggatcctacgggacttctacacacatgtgtcccgacagcagtcgtaatttgatt 601  
  
QY 481 ggaaggcctcttacttgggcatatttctccctgttcccttgccttgaatcgaatcgc 540  
DB 602 ggaaggcctcttacttgggcatatttctccctgttcccttgccttgaatcgaatcgc 661  
  
QY 541 cctcgtcttctcgtcctacccaagagaatcgtcccaactactagatgccttacaagcc 600  
DB 662 cctcgtcttctcgtcctacccaagagaatcgtcccaactactagatgccttacaagcc 721  
  
QY 601 caactcttgcacaagagcctcctcaagagcctgtgcaaccccttcccaagatgaatgag 660  
DB 722 caactcttgcacaagagcctcctcaagagcctgtgcaaccccttcccaagatgaatgag 781  
  
QY 661 ttcaattcctacagcctgtcaggtatgtc 690  
DB 781

Db 782 ttcaattctacagcctgacaggttatgtg 811

RESULT 3

AAS15360

ID AAS15360 standard; cDNA; 1475 BP.

XX AAS15360;

AC AAS15360;

DT 16-JAN-2002 (first entry)

XX cDNA encoding human PRO1356 polypeptide.

DE cDNA encoding human PRO1356 polypeptide.

XX Human; PRO1356; clone DNA64886-1601; immune-related disorder;

KW inflammatory disorder; infectious disorder; immunodeficiency disorder;

KW autoimmune disorder; renal disease; demyelinating disease; skin disease;

KW neoplasia; transplantation associated disease; gene therapy;

KW immunosuppressive; anti-inflammatory; antiasthmatic; antidiabetic; ss.

XX Homo sapiens.

OS

XX

FH Key Location/Qualifiers

FT 122..814

FT CDS /tag= a

FT /product= "PRO1356 polypeptide"

FT sig\_peptide 122..193

FT /tag= b

FT mat\_peptide 194..811

FT /tag= c

XX WO200166740-A2.

XX 13-SEP-2001.

XX 01-MAR-2001; 2001WO-US06666.

XX 03-MAR-2000; 2000US-187202P.

XX 21-MAR-2000; 2000US-191015P.

XX 30-MAY-2000; 2000WO-US14941.

XX 05-JUN-2000; 2000US-209832P.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-DEC-2000; 2000WO-US32678.

XX (GETH ) GENENTECH INC.

XX Eaton DL, Fong S, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL;

PI Tumas D, Watanabe CK, Wood WL, Zhang Z;

XX WPI; 2001-625876/72.

DR P-PSDB; AAU09178.

XX Nucleic acids encoding PRO polypeptides, useful for detecting and

PT treating immune related diseases and disorders in mammals including

PT autoimmune diseases, inflammatory diseases and asthma -

XX Claim 2; Fig 1; 122pp; English.

XX The present invention relates to the isolation of 9 novel human PRO

CC polypeptides (AAU09178-AAU09186) and the cDNA sequences encoding them.

CC The novel PRO polypeptides include PRO1356, PRO1268, PRO1884, PRO3444,

CC PRO3151, PRO4322, PRO9964, PRO10008 and PRO19598. The cDNA sequences

CC encoding these PRO polypeptides have been designated as clones

CC DNA64886-1601, DNA64903-1553, DNA84318-2520, DNA87997, DNA89273,

CC DNA92223-2567, DNA96973, DNA101921 and DNA145887 respectively.

CC Compositions (e.g. vaccines) containing PRO polypeptides and methods of

CC using these compositions are useful in the treatment and diagnosis of

CC immune-related disorders. Such disorders include immune-mediated

CC inflammatory disorders (e.g. osteoarthritis), non-immune-mediated

CC inflammatory disorders (e.g. diabetes mellitus), infectious disorders

CC (e.g. granulomatous hepatitis), immunodeficiency disorders (e.g. AIDS),

CC autoimmune disorders (e.g. rheumatoid arthritis), immune-related renal

CC diseases (e.g. cirrhosis), demyelinating diseases of the peripheral or

CC central nervous system (e.g. Guillain-Barre syndrome), immune-mediated

CC

CC skin diseases (e.g. contact dermatitis), neoplasias and transplantation

CC associated diseases. The polynucleotide sequences of the invention may

CC be used in gene therapy. AAS15360-AAS15368 represent cDNA sequences

CC encoding for the novel human PRO polypeptides of the invention.

XX

SQ Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;

Query Match 100.0%; Score 690; DB 22; Length 1475;

Best Local Similarity 100.0%; Pred. No. 7.6e-204;

Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atggcctctctggcctccaaactgtgggtgtacatctcctaggcctctctggggttttgggc 60

Db 122 atggcctctctggcctccaaactgtgggtgtacatctcctaggcctctctggggttttgggc 181

Qy 61 acactgggttgcctgctctccagctggaaacaaagtcttattgtcgggtccagcatt 120

Db 182 acactgggttgcctgctctccagctggaaacaaagtcttattgtcgggtccagcatt 241

Qy 121 gtgacagcagttggcttctccaaaggcctctgtggtgaatgtgccacacagcacagcc 180

Db 242 gtgacagcagttggcttctccaaaggcctctgtggtgaatgtgccacacagcacagcc 301

Qy 181 atcaccagtgtagcatctatagcaccctctgtggcctgcgcgtgacatccaggtgcc 240

Db 302 atcaccagtgtagcatctatagcaccctctgtggcctgcgcgtgacatccaggtgcc 361

Qy 241 caggcctatggtgacatccagctcaatctcctccctggcctgcttctctctgtgtg 300

Db 362 caggcctatggtgacatccagctcaatctcctccctggcctgcttctctctgtgtg 421

Qy 301 ggcctgagatgcacagcttctgtccaggaatcccgagccaaagacagagtgccgtagca 360

Db 422 ggcctgagatgcacagcttctgtccaggaatcccgagccaaagacagagtgccgtagca 481

Qy 361 ggtgagctcttttctcctctggagcctctcctgggtatctcctctgttgcctggaattt 420

Db 482 ggtgagctcttttctcctctggagcctctcctgggtatctcctctgttgcctggaattt 541

Qy 421 catgggattcctacgggacttctactcaccactgtgctgcctgacagatgaaattgagatt 480

Db 542 catgggattcctacgggacttctactcaccactgtgctgcctgacagatgaaattgagatt 601

Qy 481 ggagagcctcttttacttgggcattatttctcctgttctcctcctgagtggaatcattc 540

Db 602 ggagagcctcttttacttgggcattatttctcctgttctcctcctgagtggaatcattc 661

Qy 541 ctctgtcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 600

Db 662 ctctgtcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 721

Qy 601 caacctctgtccacaaggagctctccaaaggcctgtgctcaacctcccaaaagtcgaagtgag.660

Db 722 caacctctgtccacaaggagctctccaaaggcctgtgctcaacctcccaaaagtcgaagtgag 781

Qy 661 ttcaattctacagcctgacaggttatgtg 690

Db 782 ttcaattctacagcctgacaggttatgtg 811

RESULT 4

AAS21489

ID AAS21489 standard; cDNA; 1475 BP.

XX AAS21489;

AC AAS21489;

DT 24-OCT-2001 (first entry)

XX Human cDNA sequence encoding for PRO1356 polypeptide.

DE Human secretory and transmembrane; PRO; mammalian; cancer; lung;

KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;

KW

KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
KM adipocyte; A-peptide; factor VIIA; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200140466-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 01-DEC-2000; 2000MO-US32678.  
XX  
PR 01-DEC-1999; 99MO-US28301.  
PR 01-DEC-1999; 99MO-US28634.  
PR 02-DEC-1999; 99MO-US28551.  
PR 02-DEC-1999; 99MO-US28564.  
PR 02-DEC-1999; 99MO-US28565.  
PR 09-DEC-1999; 99US-0170262.  
PR 16-DEC-1999; 99MO-US30095.  
PR 20-DEC-1999; 99MO-US30911.  
PR 20-DEC-1999; 99MO-US30999.  
PR 30-DEC-1999; 99MO-US31243.  
PR 06-JAN-2000; 2000MO-US00277.  
PR 06-JAN-2000; 2000MO-US00376.  
PR 11-FEB-2000; 2000MO-US03565.  
PR 18-FEB-2000; 2000MO-US04341.  
PR 18-FEB-2000; 2000MO-US04342.  
PR 22-FEB-2000; 2000MO-US04414.  
PR 24-FEB-2000; 2000MO-US04914.  
PR 24-FEB-2000; 2000MO-US05004.  
PR 01-MAR-2000; 2000MO-US05601.  
PR 20-MAR-2000; 2000MO-US07377.  
PR 21-MAR-2000; 2000MO-US07532.  
PR 30-MAR-2000; 2000MO-US08439.  
PR 17-MAY-2000; 2000MO-US13705.  
PR 22-MAY-2000; 2000MO-US14042.  
PR 30-MAY-2000; 2000MO-US14941.  
PR 02-JUN-2000; 2000MO-US15264.  
PR 10-NOV-2000; 2000MO-US30873.  
XX  
XX (GENTECH ) GENE/TECH INC.  
XX  
PI Baker KP, Betarsini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;  
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
PI Smith V, Stewart JA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX MPI: 2001-408281/43.  
XX  
XX P-PSDB; AAU12417.  
XX  
XX Isolated, secretory and transmembrane PRO polypeptide used to detect  
XX other PRO polypeptides, link bioactive molecules to cells expressing  
XX PRO polypeptides, and detect the presence of mammalian tumours e.g.  
XX lung, breast, prostate, cervical -  
XX  
XX Claim 3; Fig 491; 813pp; English.  
XX  
XX AAS21244-AAS21518 encode for novel human secretory and transmembrane  
XX PRO polypeptides. The PRO polypeptides are useful to detect other  
XX PRO polypeptides, to link bioactive molecules to cells expressing  
XX PRO polypeptides, to modulate biological activities of cells expressing  
XX PRO polypeptides, and to detect the presence of mammalian lung, colon,  
XX breast, prostate, rectal, cervical or liver tumours by comparing PRO  
XX polypeptide expression in a cell sample to that in a control sample.  
XX Some of the 275 sequences are also useful to stimulate the release of  
XX tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
XX proliferation or differentiation of chondrocytes, the proliferation or  
XX gene expression in pericyte cells, the release of proteoglycans from  
XX cartilage, the proliferation of inner ear utricular supporting cells or  
XX of T-lymphocytes, the release of a cytokine from peripheral blood  
XX monocytes (PMNs), or the proliferation of endothelial cells. Some of  
XX the PRO polypeptides may modulate glucose or free fatty acid uptake by  
XX skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
XX to factor VIIA. The PRO polypeptides can be used in assays to identify  
XX molecules involved in binding interactions. The polynucleotides encoding

CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
CC transgenic or knock out animals and can be used in gene therapy.  
XX  
XX Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;  
SQ  
  
Query Match 100.0%; Score 690; DB 22; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.6e-204;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 atggcctcttgccctccaacttggtggctacatccctcaggctctctgggctttgggc 60  
DB |||||||  
DB 122 atggcctcttgccctccaacttggtggctacatccctcaggctctctgggctttgggc 181  
OY 61 acactggttgcacatgctctctccacagcttgaaacaagttctatgctggtccagcatt 120  
DB |||||||  
DB 182 acactggttgcacatgctctctccacagcttgaaacaagttctatgctggtccagcatt 241  
OY 121 gtgacagcagttgcttctccaaggcctctgagtgaatgtgccaacacagcagaagc 180  
DB |||||||  
DB 242 gtgacagcagttgcttctccaaggcctctgagtgaatgtgccaacacagcagaagc 301  
OY 181 atcaccacagtgtagatctatagcaaccctctgggctctccgctgacatccagctgcc 240  
DB |||||||  
DB 302 atcaccacagtgtagatctatagcaaccctctgggctctccgctgacatccagctgcc 361  
OY 241 caggccatgattgtacatccagctgacatctctccctgggctgacatctctctgtgtg 300  
DB |||||||  
DB 362 caggccatgattgtacatccagctgacatctctccctgggctgacatctctctgtgtg 421  
OY 301 ggcatgagatgacacagctctctgcaggaatcccgaccgaagacagatggtgcgttagca 360  
DB |||||||  
DB 422 ggcatgagatgacacagctctctgcaggaatcccgaccgaagacagatggtgcgttagca 481  
OY 361 ggtgagatcttttatactcttggaagcctcccgagattatccctgttgcctggaaatct 420  
DB |||||||  
DB 482 ggtgagatcttttatactcttggaagcctcccgagattatccctgttgcctggaaatct 541  
OY 421 catggatctcctacaggaacttctacccaacacagctggtccgacagatgaaattagatt 480  
DB |||||||  
DB 542 catggatctcctacaggaacttctacccaacacagctggtccgacagatgaaattagatt 601  
OY 481 ggaagagctcttacttgggacatattcttccctgttctccctgatagctggaatcattc 540  
DB |||||||  
DB 602 ggaagagctcttacttgggacatattcttccctgttctccctgatagctggaatcattc 661  
OY 541 ctctgcttctctgctcaccacagagaatgctctccaactactacgattgctccaagcc 600  
DB |||||||  
DB 662 ctctgcttctctgctcaccacagagaatgctctccaactactacgattgctccaagcc 721  
OY 601 caactcttgcacaagaagagctctccaaggtctgttaacccccaagaagtcaaggttag 660  
DB |||||||  
DB 722 caactcttgcacaagaagagctctccaaggtctgttaacccccaagaagtcaaggttag 781  
OY 661 ttcaattctacagcctacagaggtatgtg 690  
DB |||||||  
DB 782 ttcaattctacagcctacagaggtatgtg 811  
  
RESULT 5  
AAF92097  
ID AAF92097 standard; cDNA; 1475 BP.  
XX  
AC AAF92097;  
XX  
XX 15-MAY-2001 (first entry)  
XX  
XX Human PRO1356 cDNA.  
XX  
XX Human; PRO protein; mapping; ss.  
XX  
XX Homo sapiens.  
XX

PN WO200116318-A2.  
XX 08-MAR-2001.  
PD 24-AUG-2000; 2000WO-US23328.  
XX 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21090.  
PR 07-DEC-1999; 99US-0169495.  
PR 09-DEC-1999; 99US-0170262.  
PR 11-JAN-2000; 2000US-0175481.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 03-MAR-2000; 2000US-0187202.  
PR 25-APR-2000; 2000US-0199397.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 05-JUN-2000; 2000US-0209832.  
XX (GETH ) GENENTECH INC.  
PA Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;  
PI WPI: 2001-183260/18.  
DR P-PSDB; AAB87565.  
XX  
XX Eighty four nucleic acids encoding PRO polypeptides, useful in  
PT molecular biology, including use as hybridization probes, and in  
PT chromosome and gene mapping.  
XX  
XX Claim 2; Fig 79; 278pp; English.  
XX  
XX The present sequence is the coding sequence for a human PRO polypeptide  
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO  
CC antagonists or anti-PRO antibodies are useful for preparation of a  
CC medicament useful in the treatment of a condition which is responsive to  
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO  
CC protein may also be employed as molecular weight markers for protein  
CC electrophoresis. The PRO coding sequence has applications in molecular  
CC biology, including use as hybridisation probes, and in chromosome and  
CC gene mapping.  
XX  
XX Sequence 1475 BP; 345 A; 404 C; 376 G; 350 T; 0 other;  
SQ

Query Match 100.0%; Score 690; DB 22; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.6e-204;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctcttggcctccaaactgtggcctacatccctagccttctgggcttttggc 60  
DB 122 atggcctcttggcctccaaactgtggcctacatccctagccttctgggcttttggc 181  
QY 61 acactgtgtgccatgtgtctcccaagctggaacaaagtcttatgtcgtgacagcatt 120  
DB 182 acactgtgtgccatgtgtctcccaagctggaacaaagtcttatgtcgtgacagcatt 241  
QY 121 gtgacagcagtggtgtcttcccaaggcctctggtatggaatgtgccacacagcacaggc 180  
DB 242 gtgacagcagtggtgtcttcccaaggcctctggtatggaatgtgccacacagcacaggc 301  
QY 181 atcacccagtgtagcatctatagcaccttctggcctgcccctgacatccagctgcc 240  
DB 302 atcacccagtgtagcatctatagcaccttctggcctgcccctgacatccagctgcc 361  
QY 241 caggccatgatgtgacatccagtcgaatctctccctggcctgacattatctctgtgtg 300  
DB 362 caggccatgatgtgacatccagtcgaatctctccctggcctgacattatctctgtgtg 421  
QY 301 ggcagatgacagcttctctccaggaatcccgagccaaagacagagtggcggtagca 360

DB 422 ggcagatgacagctcttctgcaggaatcccgagccaaagacagagtggtggtagca 481  
QY 361 ggtgagtgcttttctcctctgagggcctcctgggattcattctgttgcctggaattctt 420  
DB 482 ggtgagtgcttttctcctctgagggcctcctgggattcattctgttgcctggaattctt 541  
QY 421 catggatcctacgggacttctactcaccactgtgctgacagcatgaaatttgagatt 480  
DB 542 catggatcctacgggacttctactcaccactgtgctgacagcatgaaatttgagatt 601  
QY 481 ggaagagctcttacttgggcattatttctcctgttctcctgtagctggaatcattc 540  
DB 602 ggaagagctcttacttgggcattatttctcctgttctcctgtagctggaatcattc 661  
QY 541 ctgtcttctctgctcatccagagaaatcgctcccaactactacgactaccacagcc 600  
DB 662 ctgtcttctctgctcatccagagaaatcgctcccaactactacgactaccacagcc 721  
QY 601 caactcttggccaaagagctctccaaaggcctgtgtaacctcccaagtcacagagtgag 660  
DB 722 caactcttggccaaagagctctccaaaggcctgtgtaacctcccaagtcacagagtgag 781  
QY 661 ttcaatctctacagcctgacaggtatgtg 690  
DB 782 ttcaatctctacagcctgacaggtatgtg 811  
RESULT 6  
AAF54296  
ID AAF54296 standard; DNA; 1475 BP.  
XX  
XX AAF54296;  
XX  
XX 02-APR-2001 (first entry)  
XX  
XX DNA encoding protein of the invention #39.  
XX  
XX Secreted; transmembrane; gene therapy; ss.  
XX  
XX Unidentified.  
XX  
XX WO200078961-A1.  
XX  
XX 28-DEC-2000.  
XX  
XX 18-FEB-2000; 2000WO-US04342.  
XX  
XX 23-JUN-1999; 99US-0141037.  
PR 20-JUL-1999; 99US-0144758.  
PR 26-JUL-1999; 99US-0145698.  
PR 01-SEP-1999; 99WO-US20111.  
PR 29-OCT-1999; 99US-0162506.  
PR 30-NOV-1999; 99WO-US28313.  
PR 02-DEC-1999; 99WO-US28351.  
PR 16-DEC-1999; 99WO-US30095.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00376.  
XX  
XX (GETH ) GENENTECH INC.  
PA Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;  
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;  
PI Watanabe CK, Williams PM, Wood WI;  
XX  
XX WPI: 2001-071395/08.  
XX  
XX Secreted and transmembrane proteins and nucleic acids designated PRO,  
PT useful as hybridization probes, in chromosome and gene mapping and gene  
PT therapy -  
XX  
XX Claim 2; Fig 77; 787pp; English.  
XX



CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.  
XX  
SQ Sequence 1481 BP; 353 A; 405 C; 375 G; 348 T; 0 other;

Query Match 100.0%; Score 690; DB 22; Length 1481;  
Best Local Similarity 100.0%; Pred. No. 7.6e-204; Mismatches 0; Indels 0; Gaps 0;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctcttggcctccaaacttggtggtacatctagggccttctgggcttttgggc 60  
Db 118 atggcctcttggcctccaaacttggtggtacatctagggccttctgggcttttgggc 177  
QY 61 acactggttgccatgctgtcccccagctggaaacaaagtcttatgtcggtccagcatt 120  
Db 178 acactggttgccatgctgtcccccagctggaaacaaagtcttatgtcggtccagcatt 237  
QY 121 gtgacagcagttggtcttccaaaggcctctggatggaatgtccacacacagcacaggc 180  
Db 238 gtgacagcagttggtcttccaaaggcctctggatggaatgtccacacacagcacaggc 297  
QY 181 atcacccagtgacatctatagcacccttctggcctgccctgacatccaggctgcc 240  
Db 298 atcacccagtgacatctatagcacccttctggcctgccctgacatccaggctgcc 357  
QY 241 caggccatgatggtgacatccagtgcaatctctccctggcctgcattatctgtggtg 300  
Db 358 caggccatgatggtgacatccagtgcaatctctccctggcctgcattatctgtggtg 417  
QY 301 ggcagtgagtgacagcttcttcgcaaggaaatcccgagcgaacagacagagtggtgagca 360  
Db 418 ggcagtgagtgacagcttcttcgcaaggaaatcccgagcgaacagacagagtggtgagca 477  
QY 361 ggtgagtgcttttctacctctgagagcctctctgggattcattctctgttgcctggaattct 420  
Db 478 ggtgagtgcttttctacctctgagagcctctctgggattcattctctgttgcctggaattct 537  
QY 421 catgggactcagggacttctactcaccactgtgctgacagcatgaaatttgagatt 480  
Db 538 catgggactcagggacttctactcaccactgtgctgacagcatgaaatttgagatt 597  
QY 481 ggagaggctcttacttggcattatttctccctgttctccctgatatgagtgaatcattc 540  
Db 598 ggagaggctcttacttggcattatttctccctgttctccctgatatgagtgaatcattc 657  
QY 541 ctctgcttttctgctcctccagagaaatcgctccaactactacgagctcctaccaagcc 600  
Db 658 ctctgcttttctgctcctccagagaaatcgctccaactactacgagctcctaccaagcc 717  
QY 601 caacctcttgccacaaggagctctccaaaggcctgggtcgaacctcccaagtcgaagttag 660  
Db 718 caacctcttgccacaaggagctctccaaaggcctgggtcgaacctcccaagtcgaagttag 777  
QY 661 tccaattcctacagcctgcaggggtatgtg 690  
Db 778 tccaattcctacagcctgcaggggtatgtg 807

## RESULT 8

ABA04424  
ID ABA04424 standard; cDNA; 1918 BP.

XX  
AC ABA04424;

XX  
DT 11-MAR-2002 (first entry)

XX  
DE Human SP82 protein encoding cDNA SEQ ID NO:1/3.

XX  
KW Human; SP82; cancer suppression; ss.

XX  
OS Homo sapiens.

XX

FH Key Location/Qualifiers  
FT CDS 520..1212  
FT /\*tag= a  
FT /product= \*SP82\*  
XX  
XX CN1313315-A.  
XX  
XX 19-SEP-2001.  
XX  
XX 13-MAR-2000; 2000CN-0111989.  
XX  
XX 13-MAR-2000; 2000CN-0111989.  
XX  
XX (SHAN-) SHANGHAI INST ONCOLOGY.  
XX  
XX Gu J, Yang S;  
XX  
XX WPI: 2002-042193/06.  
XX P-PSDB; ABB04707.  
XX  
XX New human protein able to suppress growth of cancer cells and its  
XX encoding polynucleotide sequence -  
XX  
XX Claim 5; Page 11-12 (Disclosure); 42pp; Chinese.  
XX  
XX The present sequence encodes human SP82 protein, which has cancer-  
XX suppressing activity. The present invention also describes a method  
XX for the preparation of the protein by recombination, and the application  
XX of the protein in treating diseases such as cancer.  
XX  
XX Sequence 1918 BP; 486 A; 525 C; 473 G; 434 T; 0 other;

Query Match 100.0%; Score 690; DB 24; Length 1918;  
Best Local Similarity 100.0%; Pred. No. 8.6e-204; Mismatches 0; Indels 0; Gaps 0;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctcttggcctccaaacttggtggtacatctagggccttctgggcttttgggc 60  
Db 520 atggcctcttggcctccaaacttggtggtacatctagggccttctgggcttttgggc 579  
QY 61 acactggttgccatgctgtctccagctggaaacaaagtcttatgtcggtgccaagcatt 120  
Db 580 acactggttgccatgctgtctccagctggaaacaaagtcttatgtcggtgccaagcatt 639  
QY 121 gtgacagcagtggtcttccaaaggcctctggatggaatgtccacacacagcacaggc 180  
Db 640 gtgacagcagtggtcttccaaaggcctctggatggaatgtccacacacagcacaggc 699  
QY 181 atcacccagtgacatctatagcacccttctggcctgccctgacatccaggctgcc 240  
Db 700 atcacccagtgacatctatagcacccttctggcctgccctgacatccaggctgcc 759  
QY 241 caggccatgatggtgacatccagtgcaatctctccctggcctgattatctctgtggtg 300  
Db 760 caggccatgatggtgacatccagtgcaatctctccctggcctgattatctctgtggtg 819  
QY 301 ggcagtgagtgacagcttcttgcagaaatcccgagcgaacagacagagtggtgagca 360  
Db 820 ggcagtgagtgacagcttcttgcagaaatcccgagcgaacagagagtggtgagca 879  
QY 361 ggtgagtgcttttctaccttggaggcctctctgggattcattctctgttgcctggaattct 420  
Db 880 ggtgagtgcttttctaccttggaggcctctctgggattcattctctgttgcctggaattct 939  
QY 421 catgggactcagggacttctactcaccactgttgctgacagcatgaaatttgagatt 480  
Db 940 catgggactcagggacttctactcaccactgttgctgacagcatgaaatttgagatt 999  
QY 481 ggagaggctcttacttgggattatttctccctgttctccctgatatgagtggaatcattc 540  
Db 1000 ggagaggctcttacttgggattatttctccctgttctccctgatatgagtggaatcattc 1059

OY 541 ctctgcttctctgctatcccaagaagaatcgctcccaactactactagatgctaccacgc 600  
CC |||||||  
CC ctctgcttctctgctatcccaagaagaatcgctcccaactactactagatgctaccacgc 1119  
Db 1060 ctctgcttctctgctatcccaagaagaatcgctcccaactactactagatgctaccacgc 1119  
OY 601 caacctcttgccacaagaagctctcccaagcctgtgtaacctcccaagaagtaagtgag 660  
CC |||||||  
CC caacctcttgccacaagaagctctcccaagcctgtgtaacctcccaagaagtaagtgag 1179  
Db 1120 caacctcttgccacaagaagctctcccaagcctgtgtaacctcccaagaagtaagtgag 1179  
OY 661 tcaattcctacagcctgacaggtatgtg 690  
CC |||||||  
CC tcaattcctacagcctgacaggtatgtg 1209  
Db 1180 tcaattcctacagcctgacaggtatgtg 1209

RESULT 9  
AA12585  
ID AA12585 standard; cDNA; 2742 BP.  
XX  
AC AA12585;  
XX  
DT 25-JUL-2000 (first entry)  
XX  
DE cDNA encoding a membrane associated organizational protein (HUNCT).  
XX  
KM Human: membrane associated organizational protein; HUNCT;  
KM cell proliferative disorder; cancer; autoimmune disorder;  
KM inflammatory disorder; neurological disorder; developmental disorder;  
KM vesicle trafficking; reproductive disorder; gastrointestinal disorder;  
KM renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;  
KM Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;  
KM irritable bowel syndrome; allergy; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 119..811  
FT /+tag= a  
FT /product= "membrane associated organizational protein"  
FT  
XX  
XX WO200018915-A2.  
XX  
XX PD 06-APR-2000.  
XX  
XX PF 23-SEP-1999; 99WO-US22082.  
XX  
XX PR 25-SEP-1998; 98US-0155215.  
XX PR 13-OCT-1998; 98US-0155251.  
XX PR 04-MAY-1999; 99US-0172228.  
XX  
XX PA (INCYTE) INCYTE PHARM INC.  
XX  
XX PI Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR, Lu AD, Tang YT;  
XX  
XX DR WPI: 2000-293154/25.  
XX DR P-PSDB; AA184609.  
XX  
XX PT Human membrane associated organizational protein and nucleic acid  
XX PT sequences useful in the diagnosis, treatment and prevention of cell  
XX PT proliferative associated disorders e.g. cancer, rheumatoid arthritis  
XX PT and Alzheimer's disease -  
XX  
XX PS Claim 9; Page 79-80; 84pp; English.  
XX  
XX  
XX The present sequence encodes a human membrane associated organizational  
XX protein (HUNCT). HUNCT is used for the diagnosis, treatment and  
XX prevention of cell proliferative disorders including cancer and  
XX autoimmune/inflammatory, neurological, developmental, vesicle  
XX trafficking, reproductive, gastrointestinal and renal disorders. These  
XX disorders may include atherosclerosis, leukaemia, allergies, rheumatoid  
XX arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,  
XX renal failure and irritable bowel syndrome. A vector expressing HUNCT,  
XX and an agonist of HUNCT can be used to treat or prevent a disorder  
XX associated with decreased expression or activity of HUNCT. An antagonist  
XX of HUNCT or a vector expressing the complement of a polynucleotide

CC encoding HUNCT can be used to treat or prevent a disorder associated  
CC with increased expression or activity of HUNCT. Antibodies which bind  
CC HUNCT can be used for diagnosis of disorders associated with HUNCT  
CC expression or to monitor patients being treated with HUNCT, agonists,  
CC antagonists or inhibitors of HUNCT. Assays are preferably carried out on  
CC body fluids from a patient using radioimmunoassay, enzyme linked  
CC immunosorbent assays or fluorescent activated cell sorting assays.  
CC Polynucleotides encoding HUNCT are also used in hybridisation assays to  
CC determine absence, presence or excess expression of HUNCT and to monitor  
CC regulation of HUNCT levels during disease therapy.  
XX  
XX Sequence 2742 BP; 648 A; 754 C; 682 G; 658 T; 0 other:  
SQ

Query Match 100.0%; Score 690; DB 21; Length 2742;  
Best Local Similarity 100.0%; Pred. No. 1e-203;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atggcctctctggccctcaactgtggtgctacatccttagcctctgggcttttgggc 60  
CC |||||||  
Db 119 atggcctctctggccctcaactgtggtgctacatccttagcctctgggcttttgggc 178  
OY 61 aacactggttcacatgctgtctcccccagctggaataacagtcttatgtcgtgccacatt 120  
CC |||||||  
Db 179 aacactggttcacatgctgtctcccccagctggaataacagtcttatgtcgtgccacatt 238  
OY 121 gtgaacagcagttgtgctcccaaggccctctgatatgaaatgtccacacacagcacaggc 180  
CC |||||||  
Db 239 gtgaacagcagttgtgctcccaaggccctctgatatgaaatgtccacacacagcacaggc 298  
OY 181 atcacccagttgtacatcctatagacacctctctggcctgcccgtgacatccaggctgcc 240  
CC |||||||  
Db 299 atcacccagttgtacatcctatagacacctctctggcctgcccgtgacatccaggctgcc 358  
OY 241 caggccatgattgtgtacatcccaagctgaatctcctcctgctgtcatatctctgtgtg 300  
CC |||||||  
Db 339 caggccatgattgtgtacatcccaagctgaatctcctcctgctgtcatatctctgtgtg 418  
OY 301 ggcatagatagcacagctctctccaggaatcccgagcaagaagacagatgacgtagaga 360  
CC |||||||  
Db 419 ggcatagatagcacagctctctccaggaatcccgagcaagaagacagatgacgtagaga 478  
OY 361 ggtggaagcttttcaatcctctggaggccctctggagatcaatcctctgttccctggaatct 420  
CC |||||||  
Db 479 ggtggaagcttttcaatcctctggaggccctctggagatcaatcctctgttccctggaatct 538  
OY 421 catggatcccaaggaactctcaatccacacacagtggtgcgtgacagcaatgaatttgatt 480  
CC |||||||  
Db 539 catggatcccaaggaactctcaatccacacacagtggtgcgtgacagcaatgaatttgatt 598  
OY 481 ggaaggcctcttacttgggcatcttctcctgttctcctgtaataagtggaatcattc 540  
CC |||||||  
Db 599 ggaaggcctcttacttgggcatcttctcctgttctcctgtaataagtggaatcattc 658  
OY 541 ctctgcttctctgctacatcccaagaagaatcgctcccaactactactagatgctaccacgc 600  
CC |||||||  
Db 659 ctctgcttctctgctacatcccaagaagaatcgctcccaactactactagatgctaccacgc 718  
OY 601 caacctcttgccacaagaagctctcccaagcctgtgtaacctcccaagaagtaagtgag 660  
CC |||||||  
Db 719 caacctcttgccacaagaagctctcccaagcctgtgtaacctcccaagaagtaagtgag 778  
OY 661 tcaattcctacagcctgacaggtatgtg 690  
CC |||||||  
Db 779 tcaattcctacagcctgacaggtatgtg 808

RESULT 10  
AAF93769  
ID AAF93769 standard; cDNA; 2863 BP.  
XX  
XX AAF93769;  
XX

23-MAY-2001 (first entry)  
Human cDNA encoding a membrane or secretory protein clone PSEC0059.  
Human; secretory protein; membrane protein; vaccine; gene therapy;  
rheumatoid arthritis; diabetes; ss.  
Homo sapiens.  
EP1067182-A2.  
10-JAN-2001.  
07-JUL-2000; 2000EP-0114090.  
08-JUL-1999; 99JP-0194179.  
11-JAN-2000; 2000JP-0118775.  
02-MAY-2000; 2000JP-0183766.  
(HELI-) HELIX RES INST.  
Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
WPI; 2001-093989/11.  
P-PSDB; AAB88342.  
Nucleic acids encoding secretory proteins/membrane proteins, useful in  
gene therapy or as candidate target molecules in drug development -  
Claim 1; SEQ ID 51; 609pp + CD ROM; English.  
This invention relates to nucleic acid sequences AAF93744 - AAF93916  
which encode human secretory or membrane proteins represented by  
AAB8317 - AAB88419. Included in the invention are primers  
AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
cDNA sequences of the invention. The invention also includes methods for  
the production of antibodies directed against the proteins, and cDNA  
sequences, which can be used in vaccines. The polynucleotide sequences  
can be used in gene therapy. The polynucleotide sequences and the  
proteins they encode may be used in the prevention, treatment and  
diagnosis of diseases associated with inappropriate secretory  
protein/membrane protein expression. The nucleic acids and complementary  
sequences may also be used as DNA probes in diagnostic assays  
(e.g. polymerase chain reactions (PCR)) to detect and quantify the  
presence of similar nucleic acid sequences in samples. They may also be  
used to study the expression and function of secretory proteins/membrane  
polypeptides and their role in metabolism. The polypeptides may be used  
as antigens in the production of antibodies against them and in assays to  
identify modulators (agonists and antagonists) of expression and  
activity. The antibodies and antagonists may also be used as therapeutic  
agents to down regulate expression and activity. The antibodies may also  
be used as diagnostic agents for detecting the presence of the  
polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
(ELISA). Examples of diseases which may be treated include rheumatoid  
arthritis and diabetes.  
Sequence 2863 BP; 670 A; 794 C; 711 G; 688 T; 0 other;  
Query Match 100.0%; Score 690; DB 22; Length 2863;  
Best Local Similarity 100.0%; Pred. No. 1.le-203;  
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atggcctctcttgccctccaaactgtgggctacatcctagcctctctggtcttgggc 60  
|||||  
DB 231 atggcctctcttgccctccaaactgtgggctacatcctagcctctctggtcttgggc 290  
|||||  
QY 61 acactggttgccatgctgtctcccgagctggaacaaagtctcttatgtcggtgcagcatt 120  
|||||  
DB 291 acactggttgccatgctgtctcccgagctggaacaaagtctcttatgtcggtgcagcatt 350  
|||||  
QY 121 gtgacagcaggttgcttctccaggcctctggtggaatgtgccacacagcacagggc 180  
|||||

DB 351 gtgacagcaggttgcttctccaggcctctggtggaatgtgccacacagcacagggc 410  
QY 181 atcaccagtgtagacatctatagcaccctcttggtgctgcctgacatccaggtgccc 240  
|||||  
DB 411 atcaccagtgtagacatctatagcaccctcttggtgctgcctgacatccaggtgccc 470  
|||||  
QY 241 caggccatgtagtagacatccagtgcaaatctccctccctggtgctgattctctgtggtg 300  
|||||  
DB 471 caggccatgtagtagacatccagtgcaaatctccctccctggtgctgattctctgtggtg 530  
|||||  
QY 301 ggcattgtagtagacatctcttctgcagggaatcccgagcagcaagacagagtggtgtagca 360  
|||||  
DB 531 ggcattgtagtagacatctcttctgcagggaatcccgagcagcaagacagagtggtgtagca 590  
|||||  
QY 361 ggtgagctcttttctcctctggtgagcctctggtgattctctggtgctggaattctt 420  
|||||  
DB 591 ggtgagctcttttctcctctggtgagcctctggtgattctctggtgctggaattctt 650  
|||||  
QY 421 catgggattctcagggactcttactcaccactggtgctgacagcatgaaatttgagatt 480  
|||||  
DB 651 catgggattctcagggactcttactcaccactggtgctgacagcatgaaatttgagatt 710  
|||||  
QY 481 ggaagagctcttttactggtgcatattcttctcctgttctccctgtagctggaattctc 540  
|||||  
DB 711 ggaagagctcttttactggtgcatattcttctcctgttctccctgtagctggaattctc 770  
|||||  
QY 541 ctctgcttttctgctctatccagagaaatcgctccaactactacgtacctaccagccc 600  
|||||  
DB 771 ctctgcttttctgctctatccagagaaatcgctccaactactacgtacctaccagccc 830  
|||||  
QY 601 caactcttgccaaaggagctctcctcaggcctggtgtcaacctcccaagtgcaagagtgag 660  
|||||  
DB 831 caactcttgccaaaggagctctcctcaggcctggtgtcaacctcccaagtgcaagagtgag 890  
|||||  
QY 661 ttcaattcctacagcctgacaggggtatgtg 690  
|||||  
DB 891 ttcaattcctacagcctgacaggggtatgtg 920  
|||||  
RESULT 11  
AAS16180  
ID AAS16180 standard; cDNA; 1441 BP.  
XX AAS16180;  
XX 14-FEB-2002 (first entry)  
XX Human CASB81 polynucleotide.  
XX Human; CASB81; colorectal cancer; colon-associated disease; tumour; ss;  
immunogen; cytostatic; antitumour; gene therapy.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX CDS 77..769  
XX /\*tag= a  
XX /\*product= "Human CASB81 protein"  
XX /\*transl\_except= (pos:635..637, aa:Xaa)  
XX /\*note= "Xaa= unknown"  
XX WO200180879-A2.  
XX 01-NOV-2001.  
XX 19-APR-2001; 2001WO-EP04456.  
XX 20-APR-2000; 2000GB-0009907.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Vinals De Bassols YC, Gaulis SRJ, Coche T, Orntoft T;  
XX

DR WPI: 2002-041368/05.  
P-PSDB: AAU10521.  
XX  
XX Use of a claudin 2 polypeptide and polynucleotide as vaccine for  
PT preventing and treating colorectal cancer, and as diagnostic reagents  
PT in context of the cancer  
XX  
PS Claim 8; Page 57; 68pp; English.  
XX  
XX The invention relates to the use of CASB81 polypeptides and  
CC polynucleotides for the manufacture of a medicament for treating a  
CC patient suffering from or susceptible to colorectal cancer or other  
CC colon-associated tumours or diseases. The sequences act as agents for  
CC specific prophylactic or therapeutic immunisation against tumours because  
CC they are specifically expressed or highly over-expressed by  
CC tumours compared to normal cells, and can thus be targeted by  
CC antigen-specific immune mechanisms leading to the destruction of tumour  
CC cells. They can be used in development of an antibody immunospecific for  
CC CASB81 protein, which can be used for diagnosis of tumour cells. Peptides  
CC incorporating epitopes of CASB81 provide suitable immunogens. Presence of  
CC or susceptibility to colorectal cancer can be diagnosed by analysing the  
CC presence or amount of CASB81 DNA expression or CASB81 protein activity.  
CC This sequence represents a cDNA encoding a human CASB81 polypeptide.  
XX  
SQ Sequence 1441 BP; 337 A; 399 C; 357 G; 345 T; 3 other:

Query Match 99.9%; Score 689.6; DB 24; Length 1441;  
Best Local Similarity 99.9%; Pred. No. 9.9e-204;  
Matches 689; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctcaactgtgggtacacatcctagcctcttgaggcctttgggc 60  
DB 77 atggcctctcttgccctcaactgtgggtacacatcctagcctcttgaggcctttgggc 136  
QY 61 acactgttgcacatgcgtctcccaagctggaaacaaattctatgtggtgccagcatt 120  
DB 137 acactgttgcacatgcgtctcccaagctggaaacaaattctatgtggtgccagcatt 196  
QY 121 gtgacacagcttggtctctcccaagcctctggatgtgacatgtgacacacacagcagc 180  
DB 197 gtgacacagcttggtctctcccaagcctctggatgtgacatgtgacacacagcagc 256  
QY 181 atcacccagtgtagacatctatagaccctcttggtcctgcgcgtgacatccagcctgcc 240  
DB 257 atcacccagtgtagacatctatagaccctcttggtcctgcgcgtgacatccagcctgcc 316  
QY 241 caggccatgatgtgacatccatcagtgcaatctccctcctggtcctgatatctctgtgtg 300  
DB 317 caggccatgatgtgacatccatcagtgcaatctccctcctggtcctgatatctctgtgtg 376  
QY 301 ggcctgagatgacagctctctgcaggaatcccgagccaaagacagatggcggtagca 360  
DB 377 ggcctgagatgacagctctctgcaggaatcccgagccaaagacagatggcggtagca 436  
QY 361 ggtgagatctttcatcctcttgaggcctcctggtgattcattcctgtgctgagatctt 420  
DB 437 ggtgagatctttcatcctcttgaggcctcctggtgattcattcctgtgctgagatctt 496  
QY 421 catggatctctacggagcttactaccacactgtgtcctgacagcaatgaattttagatt 480  
DB 497 catggatctctacggagcttactaccacactgtgtcctgacagcaatgaattttagatt 556  
QY 481 gggagagctcttactctgggcatatttctcctgttctcctcctgattgctggaatctatc 540  
DB 557 gggagagctcttactctgggcatatttctcctgttctcctcctgattgctggaatctatc 616  
QY 541 ctctgctttcctcctatcccaagagaatgcctccaactactagatgcttaccacagcc 600  
DB 617 ctctgctttcctcctatcccaagagaatgcctccaactactagatgcttaccacagcc 676  
QY 601 caacctctgccaagaagagctctccaaggtctgttcaacctcccaagaatgaagattgag 660  
|||||

DB 677 caacctctgccaagaagagctctccaaggtctgttcaacctcccaagaatgaagattgag 736  
QY 661 ttcaattcctacagcctgacaggtatgtg 690  
DB 737 ttcaattcctacagcctgacaggtatgtg 766  
|||||

## RESULT 12

AA97818  
ID AA97818 standard; cDNA; 1524 BP.

AC AA97818;

XX 23-SEP-1999 (first entry)

DE Human secreted protein encoding cDNA #6.

KW Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;  
diagnostic; gene therapy; chromosome mapping; secretion vector; ss.

OS Homo sapiens.

PN W09925825-A2.

XX 27-MAY-1999.

PD 13-NOV-1998; 98WO-IB01862.

XX 04-SEP-1998; 98US-0099273.

PR 13-NOV-1997; 97US-0066677.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

XX (GEST ) GENSET.

PA Bouquellet L, Duclert A, Dumas Milne Edwards J;

PI WPI: 1999-147472/29.

DR P-PSDB; AA936134.

XX Extended cDNAs encoding secreted proteins

PT Example 28; Page 155-156; 307pp; English.

PS AAX97813-X97906 represent extended cDNA's which encode novel human

CC secreted proteins (see AAY36129-Y36222) and which have cytosolic,

CC thrombotic and osteopathic activity. The extended cDNAs can be used to

CC express secreted proteins or parts of them or to obtain antibodies

CC capable of binding to the secreted proteins. They may also be used in

CC diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC Uses also include design of expression vectors and secretion vectors.

XX Sequence 1524 BP; 365 A; 404 C; 382 G; 353 T; 20 other:

Query Match 99.9%; Score 689.2; DB 20; Length 1524;  
Best Local Similarity 99.7%; Pred. No. 1.4e-203;  
Matches 688; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcctctcttgccctcaactgtgggtacacatcctagcctcttgaggcctttgggc 60  
DB 160 atggcctctcttgccctcaactgtgggtacacatcctagcctcttgaggcctttgggc 219  
QY 61 acactgttgcacatgcgtctcccaagctggaaacaaattctatgtcgtgacagcatt 120  
DB 220 acactgttgcacatgcgtctcccaagctggaaacaaattctatgtcgtgacagcatt 279  
QY 121 gtgacacagcttggtctctcccaaggtcctctgtagtgaattgtgacacacagcagcagc 180  
DB 280 gtgacacagcttggtctctcccaaggtcctctgtagtgaattgtgacacacagcagcagc 339  
|||||

```
QY 181 atcaccagtgatgacatctatagcaccccttctggcctgccgctgacatccaggctgcc 240
    |||||||
Db 340 atcaccagtgatgacatctatagcaccccttctggcctgccgctgacatccaggctgcc 399
    |||||||
QY 241 caggccatgatgtagacatccagtgcaatctctccctggcctgacatctctctgtgtg 300
    |||||||
Db 400 caggccatgatgtagacatccagtgcaatctctccctggcctgacatctctctgtgtg 459
    |||||||
QY 301 ggcagatgacagatctcttccaggaatcccgagccaaagacagagtggggtgagca 360
    |||||||
Db 460 ggcagatgacagatctcttccaggaatcccgagccaaagacagagtggggtgagca 519
    |||||||
QY 361 ggtggagcttttttcatctcttgaggcctctctgggattctctgtggaattctt 420
    |||||||
Db 520 ggtggagcttttttcatctcttgaggcctctctgggattctctgtggaattctt 579
    |||||||
QY 421 catggatctcagggacttctactcaccactggtgctgacagcatgaaatttgagatt 480
    |||||||
Db 580 catggatctcagggacttctactcaccactggtgctgacagcatgaaatttgagatt 639
    |||||||
QY 481 gggagagctcttacttgggcatatttcttccctgttctccctgacagtggaatcatc 540
    |||||||
Db 640 gggagagctcttacttgggcatatttcttccctgttctccctgacagtggaatcatc 699
    |||||||
QY 541 ctctgttttctgctcatcccgagaaatcgctccaactactacgactgctaccagcc 600
    |||||||
Db 700 ctctgttttctgctcatcccgagaaatcgctccaactactacgactgctaccagcc 759
    |||||||
QY 601 caacctgtgccaaaggagctctccaaggcctgggtcaacctcccaagtcaggatgag 660
    |||||||
Db 760 caacctgtgccaaaggagctctccaaggcctgggtcaacctcccaagtcaggatgag 819
    |||||||
QY 661 ttcaattctacagctgacagggtatgtg 690
    |||||||
Db 820 ttcaattctacagctgacagggtatgtg 849
    |||||||

RESULT 13
AAI58013
ID AAI58013 standard; cDNA; 1540 BP.
XX
AC AAI58013;
XX
XX
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 216.
XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WQ200153312-A1.
XX
PD 26-JUL-2001.
XX
XX
PF 26-DEC-2000; 2000WQ-US34263.
XX
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX
(PHYSE-) HYSEQ INC.
XX
```

```
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM38857.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 216; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1540 BP; 365 A; 419 C; 389 G; 367 T; 0 other;
```

```
Query Match 99.8%; Score 688.4; DB 22; Length 1540;
Best Local Similarity 99.9%; Pred. No. 2.4e-203;
Matches 689; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggcctcttggcctccaaacttggtgggtacatctcctaggccttctgggcttttgggc 60
    |||||||
Db 183 atggcctcttggcctccaaacttggtgggtacatctcctaggccttctgggcttttgggc 242
    |||||||
QY 61 acactggttgccatgctgctcccaagctggaaacaagtcttattgtcgtgccaagatt 120
    |||||||
Db 243 acactggttgccatgctgctcccaagctggaaacaagtcttattgtcgtgccaagatt 302
    |||||||
QY 121 gtgacagcagttgcttctccaaggcctctggtggaatgtgccaacacacagcagcgc 180
    |||||||
Db 303 gtgacagcagttgcttctccaaggcctctggtggaatgtgccaacacacagcagcgc 362
    |||||||
QY 181 atcacccagtgacatctatagcaccccttctggcctgccgcgtgacatccaggctgcc 240
    |||||||
Db 363 atcacccagtgacatctatagcaccccttctggcctgccgcgtgacatccaggctgcc 422
    |||||||
QY 241 caggccatgagtgtagacatccagtgcaatctctcctcctggcctgacatctctctgtgtg 300
    |||||||
Db 423 caggccatgagtgtagacatccagtgcaatctctcctcctggcctgacatctctctgtgtg 482
    |||||||
QY 301 ggcagatgacacagcttcttccagaaatcccgagccaaagacagagtgccgtgagca 360
    |||||||
Db 483 ggcagatgacacagcttcttccagaaatcccgagccaaagacagagtgccgtgagca 542
    |||||||
QY 361 ggtggagcttttttcatctcttgaggcctcctgggattcattctctgttgcctggaattct 420
    |||||||
Db 543 ggtggagcttttttcatctcttgaggcctcctgggattcattctctgttgcctggaattct 602
    |||||||
QY 421 catggatctcagggacttctactcaccactggtgcttgacagcatgaaatttgagatt 480
    |||||||
Db 603 catggatctcagggacttctactcaccactggtgcttgacagcatgaaatttgagatt 662
    |||||||
QY 481 gggagagctcttacttgggcaattatttctcctgttctcctgatagctggaatcatc 540
    |||||||
Db 663 gggagagctcttacttgggcaattatttctcctgttctcctgatagctggaatcatc 722
    |||||||
QY 541 ctctgttttctcctgctcatcccgagaaatcgctccaactactacgactgctaccagcc 600
```

|||||  
Db 723 ctctgcttttctcgtcctatgcacagagaatgcctccaactactatgactgcttaccagcc 782  
OY 601 caacctctgcccacaagagagctctccaagcgctggtcaccctcccaagaagtagag 660  
Db 783 caacctctgcccacaagagagctctccaagcgctggtcaccctcccaagaagtagag 842  
OY 661 ttcaattcctacagcctgcacaggtatgtg 690  
Db 843 ttcaattcctacagcctgcacaggtatgtg 872  
RESULT 14  
AA159799/C  
ID AA159799 standard; cDNA; 1761 BP.  
AC AA159799;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 3788.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000MO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSFO INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RJ;  
XX  
DR WPI, 2001-442253/47.  
DR P-PSDB; AAM40643.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 3788; 10078bp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and

CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1761 BP; 407 A; 446 C; 478 G; 430 T; 0 other;  
Query Match 99.8%; Score 688.4; DB 22; Length 1761;  
Best Local Similarity 99.9%; Pred. NO. 2.6e-203;  
Matches 689; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 atggcctccttggtcccaactgtgtggtctacatccctcctccttggtggtcttttggc 60  
Db 1641 ATGGCCTCTTGGCTCCCAACTTGGGCTACATCCATGAGGCTTCTGGGCTTTGGGC 1582  
OY 61 aacatggtcgcctgctgctcccaagctggaacaagctctatgttggtccagcatt 120  
Db 1581 ACATGTTGGCTATGCTGCTCCCAAGTGAACAAAGTTTATGTGCGTCCAGCATTT 1522  
OY 121 gtgaagcagctgtgctctcctcaaggcctctgtagatgtagtgcacacacagcagcg 180  
Db 1521 GTGACAGCAGTTGGCTTTCCCAAGGCGCTGTGATGGAATGTGCCACACAGCAGCAG 1462  
OY 181 atcaccacagctgtgaacatctatagacacctctggtgctgcccgtggaatccagctgc 240  
Db 1461 ATCACCAAGTGTGACATCTATAGCACACCCCTTGGGCTGCCCGCTGCATCCAGCTGCC 1402  
OY 241 caggcagatgtagtgcacacagctgcacatctcctcctgctgctgctatctctgtgtr 300  
Db 1401 CAGGCCATGATGTGATGATCATTCAGTGCATCTCTCTCCCTGGCTTATCTGTGTGTG 1342  
OY 301 ggcataagatgcacagctctctgcacaggaatccagcccaagaagacagatggtcgtagca 360  
Db 1341 GGCATGAGATGCACAGTCTTTCGCCAGGAATCCCGAGCCAAAGACAGATGGCGGATGCA 1282  
OY 361 ggtggaagcttttcaatcccttggaagcctcctggaattcaatcctgtgctggaatcct 420  
Db 1281 GGTGAGATCTTTTTCATCTCTGGAGCCCTCCGGATTCATTCCTGTCCTGGAAATCTT 1222  
OY 421 catggatcctcctcaggaactctactcaccacatggtgcctggaagatggaattgagatt 480  
Db 1221 CATGGATCTCTACGGGACTTCTACTACCATGTCCTGACACGATGAATTTAGATT 1162  
OY 481 ggaagagctccttacttggtgcatattcttcctcgttctcctcatagatctggaatcac 540  
Db 1161 GGAGAGGCTCTTACTTGGGCAATTTCTCTCCCTGTTCTCCCTGATAGCTGAATTCATC 1102  
OY 541 ctctgcttttctgctcatcccaagagaatcgctccactactacgtagcttaccagcc 600  
Db 1101 CTTGCTTTTCTCTGCTCATGCGAGAAATCGCTCCAACTACTACGATGCTTACCAAGCC 1042  
OY 601 caacctctgcccacaagagagctctccaagcgctggtcaccctcccaagaagtagag 660  
Db 1041 CAACCTTTGGCCACAAGGAGCTCTCAAGCGCTGTCGTAACCTCCCAAGTCAAGGTGAG 982  
OY 661 ttcaattcctacagcctgcacaggtatgtg 690  
Db 981 TTCAATTCTTACAGCCTGCACAGGATATGTG 952  
RESULT 15  
AAS16181  
ID AAS16181 standard; DNA; 779 BP.  
XX  
AC AAS16181;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
DE Murine CASB81 polynucleotide.  
XX  
KW Mouse; CASB81; colorectal cancer; colon-associated disease; tumour; ds;  
KW immunogen; cytostatic; antitumour; gene therapy.  
XX

OS	Mus musculus.
XX	
FX	
FH	Key Location/Qualifiers
FT	48..740
FT	/**tag= a
FT	/product= "Murine CASB81 protein"
XX	
PN	WO200180879-A2.
XX	
PD	01-NOV-2001.
XX	
PF	19-APR-2001; 2001WO-EP04456.
XX	
PR	20-APR-2000; 2000GB-0009907.
XX	
PA	(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX	
PI	Vinals De Bassols YC, Gaulis SRJ, Coche T, Orntoft T;
XX	
DR	WPI; 2002-041368/05.
XX	
PS	P-PSDB; AAU10522.
XX	
PT	Use of a claudin 2 polypeptide and polynucleotide as vaccine for
PT	preventing and treating colorectal cancer, and as diagnostic reagents
PT	in context of the cancer -
XX	
XX	Claim 4; Page 57-58; 68pp; English.
XX	
CC	The invention relates to the use of CASB81 polypeptides and
CC	polynucleotides for the manufacture of a medicament for treating a
CC	patient suffering from or susceptible to colorectal cancer or other
CC	colon-associated tumours or diseases. The sequences act as agents for
CC	specific prophylactic or therapeutic immunisation against tumours because
CC	they are specifically expressed or highly over-expressed in colorectal
CC	tumours compared to normal cells, and can thus be targeted by
CC	antigen-specific immune mechanisms leading to the destruction of tumour
CC	cells. They can be used in development of an antibody immunospecific for
CC	CASB81 protein, which can be used for diagnosis of tumour cells. Peptides
CC	incorporating epitopes of CASB81 provide suitable immunogens. Presence of
CC	or susceptibility to colorectal cancer can be diagnosed by analysing the
CC	presence or amount of CASB81 DNA expression or CASB81 protein activity.
CC	This sequence represents DNA encoding a mouse CASB81 polypeptide.
XX	
SQ	Sequence 779 BP; 152 A; 205 C; 216 G; 206 T; 0 other;
	Query Match 79.1%; Score 546; DB 24; Length 779;
	Best Local Similarity 87.0%; Pred. No. 3e-159;
	Matches 600; Conservative 0; Mismatches 90; Indels 0; Gaps 0
Qy	1 atggcctctcttggcctccaactgtggtgcacatcctagccctctctgtggcgcttttgaggc 60
Db	48 atggcctcccttggcgttcactggtgggtcacatcctagcccttttgggctgttaggc 107
Qy	61 acactggtgccatgctgctcccagctcggaaacaagtcttatgtcgggtccagcatt 120
Db	108 acatccattgccatgctgtcttcccccaactggcggaacgagttcctatgttgtgccagcatt 167
Qy	121 gtgcacagcagttgcttctccaaggcctctgcatggaatgtgcacacacagcagcaggc 180
Db	168 gtgcgcgcggttggtctttccaaggcctctgtgatggagttgtgcacacagcagcaggc 227
Qy	181 atcaccacagtgtgacatctatagcacctctctggcgctgcccgctgacatccagctgcc 240
Db	228 atcaccacagtcgatatactacagtaaccttttaggaacttctgctgacatccagcgtgcc 287
Qy	241 caggccatgatggtgacatccagtcgaatctctccctggcctggaattatctctgtggtg 300
Db	288 caggccatgatggtgacgtccagtcgaatgtcctgcgtggtgttatctatctctgtggtg 347
Qy	301 ggcattgagatgcacagcttcttctgcaggaaatcccgagccaaagacagagtggtgcgtaga 360
Db	348 ggcatgagatgcacagctgtcttgcaggatctctgagcttaagacagagtggtgtgtagtg 407



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2002, 18:58:54 ; Search time 108.28 Seconds

(without alignments)  
163.332 Million cell updates/sec

Title: US-09-663-600A-139\_COPY\_36\_107

Perfect score: 1 atgcctctctgtccctcca.....tttgggcacactgttgc 72

Sequence: IDENTITY\_NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28	38.9	4046	1	US-07-793-961A-1
C 2	28	38.9	4046	1	US-08-240-357-1
C 3	25.6	35.6	2104	1	US-07-984-044A-4
C 4	25.6	35.6	2104	1	US-08-458-393-4
C 5	25.6	35.6	9551	1	US-08-056-200-93
C 6	25.6	35.6	9551	2	US-08-600-644-93
C 7	24.6	34.2	700	4	US-08-998-416-1131
C 8	24.4	33.9	1584	1	US-08-361-920-24
C 9	24.4	33.9	1584	1	US-08-479-939-24
C 10	24.4	33.9	1584	1	US-08-483-432-24
C 11	24.4	33.9	4919	4	US-08-456-200B-2
C 12	24.4	33.9	7032	2	US-08-149-097D-24
C 13	24.4	33.9	7032	3	US-08-949-386-24
C 14	24.4	33.9	7032	3	US-08-450-562-24
C 15	24.4	33.9	7032	3	US-08-984-709A-24
C 16	24.4	33.9	7089	3	US-08-949-386-25
C 17	24.4	33.9	7089	3	US-08-450-562-25
C 18	24.4	33.9	7089	3	US-08-984-709A-25
C 19	24.2	33.6	2226	3	US-08-889-425-1
C 20	23.6	32.8	734	4	US-09-332-191-16
C 21	23.6	32.8	734	4	US-09-332-200-16
C 22	23.6	32.8	734	4	US-09-332-197-16
C 23	23.6	32.8	1896	4	US-09-232-201-16
C 24	23.6	32.8	1896	4	US-09-232-200-30
C 25	23.6	32.8	1896	4	US-09-232-197-30
C 26	23.6	32.8	1896	4	US-09-232-201-30
C 27	23.6	32.8	2907	4	US-09-232-200-52

C 28	23.6	32.8	2907	4	US-09-232-197-52	Sequence 52, Appl
C 29	23.6	32.8	2907	4	US-09-232-201-52	Sequence 52, Appl
C 30	23.6	32.8	2917	4	US-09-232-200-26	Sequence 26, Appl
C 31	23.6	32.8	2917	4	US-09-232-197-26	Sequence 26, Appl
C 32	23.6	32.8	2917	4	US-09-232-201-26	Sequence 26, Appl
C 33	23.4	32.5	2343	2	US-09-018-760-3	Sequence 3, Appl
C 34	23.4	32.5	4131	3	US-08-726-214-11	Sequence 11, Appl
C 35	23.2	32.2	1516	3	US-09-059-522-4	Sequence 4, Appl
C 36	23.2	32.2	1516	3	US-09-382-027-2	Sequence 2, Appl
C 37	23.2	32.2	2057	3	US-09-059-522-2	Sequence 2, Appl
C 38	23.2	32.2	2057	3	US-09-382-027-2	Sequence 2, Appl
C 39	23	31.9	2384	1	US-07-814-964-10	Sequence 10, Appl
C 40	23	31.9	2384	1	US-08-258-442-10	Sequence 10, Appl
C 41	23	31.9	2384	1	US-08-328-809-5	Sequence 5, Appl
C 42	23	31.9	2384	1	PCT-US92-11107-10	Sequence 10, Appl
C 43	23	31.9	4821	1	US-08-053-614-3	Sequence 3, Appl
C 44	23	31.9	4821	1	US-08-316-397B-3	Sequence 3, Appl
C 45	23	31.9	4821	2	US-09-034-306-3	Sequence 3, Appl

## ALIGNMENTS

```
RESULT 1
US-07-793-961A-1/c
; Sequence 1, Application US/0793961A
; Patent No. 5334521
; GENERAL INFORMATION:
; APPLICANT: Yoshihiro Iehikawa
; TITLE OF INVENTION: Cloning and Character-
; TITLE OF INVENTION: Iization of a Cardiac Adenyllyl Cyclase
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alan M. Gordon
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street,
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII from DM4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/793.961A
; FILING DATE: 19911118
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2719
; TELEFAX: 203 321 2971
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs listed
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-793-961A-1
; Query Match 38.9%; Score 28; DB 1; Length 4046;
; Best Local Similarity 66.7%; Pred. No. 0.57;
```



COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/056,200  
 FILING DATE: 30-APR-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fedrick, Michael F.  
 REGISTRATION NUMBER: 36,799  
 REFERENCE/DOCKET NUMBER: NIH054.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (714) 760-0404  
 TELEFAX: (714) 760-9502  
 INFORMATION FOR SEQ ID NO: 93:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9551 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO

```

1  RESULT 6
2  US-08-800-644-93/c
3  ; Sequence 93, Application US/08800644
4  ; Patent No. 5958752
5  ;
6  ; GENERAL INFORMATION:
7  ; APPLICANT: Steinert, Peter M.
8  ; APPLICANT: Lee, Seung-Chul
9  ; APPLICANT: Kim, In-Gyu
10 ; APPLICANT: Chung, Soo-Il
11 ; APPLICANT: Park, Sang-Chul
12 ;
13 ; TITLE OF INVENTION: Trichohyalin
14 ; TITLE OF INVENTION: Methods of us
15 ; NUMBER OF SEQUENCES: 117
16 ;
17 ; CORRESPONDENCE ADDRESS:
18 ; ADDRESSEE: Knobbe, Martens, Ol
19 ; STREET: 620 Newport Center Dri
20 ; CITY: Newport Beach
21 ; STATE: CA
22 ; COUNTRY: U.S.A.
23 ; ZIP: 92660
24 ;
25 ; COMPUTER READABLE FORM:
26 ; MEDIUM TYPE: Floppy disk
27 ; COMPUTER: IBM PC compatible.
28 ; OPERATING SYSTEM: PC-DOS/MS-D
29 ; SOFTWARE: PatentIn release #1.
30 ;
31 ; CURRENT APPLICATION DATA:
32 ; APPLICATION NUMBER: US/08/800,
33 ; FILING DATE: 14-FEB-1997
34 ; CLASSIFICATION: 424
35 ; PRIOR APPLICATION DATA:
36 ; APPLICATION NUMBER: US 08/056,
37 ; FILING DATE: 30-APR-1993
38 ;
39 ; ATTORNEY/AGENT INFORMATION:
40 ; NAME: Fedrick, Michael F.
41 ; REGISTRATION NUMBER: 36,799
42 ; REFERENCE/DOCKET NUMBER: NIH05
43 ;
44 ; TELECOMMUNICATION INFORMATION:
45 ; TELEPHONE: (714) 760-0404
46 ; TELEFAX: (714) 760-9502
47 ;
48 ; INFORMATION FOR SEQ ID NO: 93:
49 ; SEQUENCE CHARACTERISTICS:
50 ; LENGTH: 951 base pairs
51 ; TYPE: nucleic acid
52 ; STRANDEDNESS: single
53 ; TOPOLOGY: linear
54 ; MOLECULE TYPE: cDNA
55 ; HYPOTHETICAL: NO
56 ; ANTI-SENSE: NO

```

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-800-644-93

Query Match 35.6% Score 25.6; DB 2; Length 9551;  
Best Local Similarity 62.5% Pred. No. 4.8;  
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 tggcctctcttggtcccaactgttggtctacatcctagagcctctggtctttggca 61  
DB 5021 TGTGCGCTCTCTCCGATGAGTCCTCTGAGGCGCTCTCTCTCTCTGAGAGCTGTGGCA 4962  
QY 62 cac 65  
DB 4961 CGCT 4958

RESULT 7  
US-08-998-416-1131/c  
Sequence 1131, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38, 241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1131:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:  
ORGANISM: P61689RP  
US-08-998-416-1131

Query Match 34.2% Score 24.6; DB 4; Length 700;  
Best Local Similarity 61.9% Pred. No. 5.7;  
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 tggcctctcttggtcccaactgttggtctacatcctagagcctctggtctttggca 61  
DB 609 TGGACGGGCTTGTTCATGAACTTGTGCTCACGTCACACACCTCTTGGCATGTGGCA 550  
QY 62 cac 64  
DB 549 GAC 547

RESULT 8  
US-08-361-920-24/c  
Sequence 24, Application US/08361920  
Patent No. 5457046  
GENERAL INFORMATION:  
APPLICANT: Woeldike, Helle F.  
APPLICANT: Hagen, Frederick  
APPLICANT: Hjort, Carsten M.  
APPLICANT: Sven, Hastup  
TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
TITLE OF INVENTION: or Hemicellulose  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5457046 No. 5457046disk of No. 5457046th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,920  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/940,860  
FILING DATE: 28-OCT-1992  
APPLICATION NUMBER: DK 1158/90  
FILING DATE: 09-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00124  
FILING DATE: 08-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3435, 204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0298  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1584 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Fusarium oxysporum  
FEATURE:  
NAME/KEY: CDS



Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 4 gctctcttgctccacttggtggctctacatcctagcctcttggtggcga 61  
db 162 GCAGCGTAGTCTCGAGCTTGGGGTGCTCTCTTGGCTTGGCGGGGTTGAGCA 105

## RESULT 11

US-08-456-200B-2  
; Sequence 2, Application US/08456200B  
; Patent No. 6229000  
; GENERAL INFORMATION:  
; APPLICANT: Franz, Jurgen; Weingartner, Bernhard;  
; APPLICANT: Unterdeck, Axel; Rae, Peter  
; TITLE OF INVENTION: TISSUE-SPECIFIC HUMAN NEURONAL  
; TITLE OF INVENTION: CALCIUM CHANNEL SUB-TYPES AND  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS  
; STREET: 660 White Plains Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10591-5144  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB  
; MEDIUM TYPE: Storage  
; COMPUTER: NEC Powermate SX/20  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,200B  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/094,712  
; FILING DATE: 19-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/858,278  
; FILING DATE: 26-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/064,778  
; FILING DATE: 19-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 41 10 785  
; FILING DATE: 04-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kuhl, G. Biscoe  
; REGISTRATION NUMBER: 33,141  
; REFERENCE/DOCKET NUMBER: Bayer 8398.3-KGB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914) 332-1700  
; TELEFAX: (914) 332-1844  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4919 nucleotides  
; TYPE: Nucleotide  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; MOLECULE TYPE: cDNA  
; US-08-456-200B-2

Query Match 33.9%; Score 24.4; DB 4; Length 4919;  
Best local Similarity 63.8%; Pctd. No. 11;  
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 15 cctccacttggtggctctacatcctagcctcttggtggcga 72  
db 2566 CTTCCGAGACTGTGGAACATCTGAGCTTGTGTGCTGCTGCGCATGTGTGCC 2623

## RESULT 12

US-08-149-097D-24  
; Sequence 24, Application US/08149097D  
; Patent No. 5874236  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael  
; APPLICANT: Ellis, Steven  
; APPLICANT: Williams, Mark  
; APPLICANT: Feldman, Daniel  
; APPLICANT: McCue, Ann  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/149,097D  
; FILING DATE: 05-NOV-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/105,536  
; FILING DATE: 11-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US92/06903  
; FILING DATE: 14-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/914,231  
; FILING DATE: 13-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/868,354  
; FILING DATE: 10-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/745,206  
; FILING DATE: 15-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/620,250  
; FILING DATE: 30-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/482,384  
; FILING DATE: 20-FEB-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/603,751  
; FILING DATE: 04-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US89/01408  
; FILING DATE: 04-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/176,899  
; FILING DATE: 04-APR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-55038  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 238-0999  
; TELEFAX: (619) 238-0062  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7032 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

```

; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 169..6921
; OTHER INFORMATION: /product= "AlphaE-1 subunit of
; OTHER INFORMATION: human calcium channel"
; US-08-149-097D-24

Query Match 33.9%; Score 24.4; DB 2; Length 7032;
Best Local Similarity 63.8%; Pred. No. 12;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 15 cctccaaattgtggctacatcctagcctctctgggctttttgggcacactggtggcc 72
Db 3759 CTTCGAGACTTGTGGACATCCTTGGACTTGTGTGTCGTTGGCGCATTTGGTGCC 3816

RESULT 13
US-08-949-386-24
; Sequence 24, Application US/08949386
; Patent No. 6090623
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Gillespie, Allison
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: US
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/949,386
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,012
; FILING DATE: 11-AUG-1994
; APPLICATION NUMBER: 08/149,097
; FILING DATE: 5-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 519808
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7032 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..6921

```

```

1 PRIOR APPLICATION DATA: 07/868,354
2 APPLICATION NUMBER: 07/868,354
3 FILING DATE: 10-APR-1992
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: PCT/US92/06903
6 FILING DATE: 14-AUG-1992
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: 07/745,206
9 FILING DATE: 15-AUG-1991
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: 07/620,250
12 FILING DATE: 30-NOV-1990
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: 07/603,751
15 FILING DATE: 08-NOV-1990
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: 07/482,384
18 FILING DATE: 02-FEB-1990
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: PCT/US89/01408
21 FILING DATE: 04-APR-1989
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 07/176,899
24 FILING DATE: 04-APR-1988
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Seidman, Stephanie L.
27 REGISTRATION NUMBER: 33,779
28 REFERENCE/DOCKET NUMBER: 6362-519812
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (619) 238-0999
31 TELEFAX: (619) 238-0062
32 INFORMATION FOR SEQ ID NO: 24:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 7032 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: double
37 TOPOLOGY: linear
38 MOLECULE TYPE: DNA (genomic)
39 FEATURE:
40 NAME/KEY: CDS
41 LOCATION: 166..6921
42 OTHER INFORMATION: /standard_name="Alpha-1E-1"
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
93
```

```

Query Match      33.9%: Score 24.4: DB 3, Length 7032:
Best Local Similarity 63.8%: Pred. No. 12:
Matches 37: Conservative 0: Mismatches 21: Indels 0: Gaps 0:

OY      15  cctccactgtggagctacatcctatgacctctgtggagctttgggacacgtgtgcc 72
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3759  cttccgacagcttggcgaacatcttcgactttttgtgtgtgttcgtggccgcatyttgtgtgcc 3816

RESULT 15
US-08-984-7094-24
: Sequence 24, Application US/08984709A
: Patent No. 6320032 .
:
: GENERAL INFORMATION:
: APPLICANT: Williams, Mark E.
: APPLICANT: Stauderman, Kenneth A.
: APPLICANT: Harpold, Michael M.
: TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
: TITLE OF INVENTION: METHODS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Heller Ehtman White & McAniffre
: STREET: 4250 Executive Square, Suite 700
: CITY: La Jolla
: STATE: California
: COUNTRY: US
: ZIP: 92037
:
: COMPUTER READABLE FORM:

```

```

1      MEDIUM TYPE: Diskette
2      COMPUTER: IBM Compatible
3      OPERATING SYSTEM: DOS
4      SOFTWARE: FASTSEQ version 1.5
5
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/08/984,709A
8      FILING DATE: 02-DEC-1997
9      CLASSIFICATION: 435
10
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Seidman, Stephanie L.
13     REGISTRATION NUMBER: 33,779
14     REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)
15
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE: (619) 450-8400
18     TELEFAX: (619) 587-5360
19
20     INFORMATION FOR SEQ ID NO: 24:
21
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 7032 base pairs
24     TYPE: nucleic acid
25     STRANDEDNESS: double
26
27     TOPOLOGY: linear
28
29     MOLECULE TYPE: DNA (genomic)
30
31     FEATURE:
32     NAME/KEY: CDS
33     LOCATION: 166..6921
34
35     OTHER INFORMATION: /standard_name="Alpha-1E-1"
36
37 US-08-984-709A-24

```

Query Match	33.9%	Score 24.4	DB 4	Length 7032
Best Local Similarity	63.8%	Pred. No. 12		
Matches	37	Conservative	0	Mismatches 21
			Indels	0
			Gaps	0
Qy	15	ccctcaacttctgtgactacatccctctagagccctctctgtgagcttttgggcacacgtgtgc	72	
Db	3759	cttccgagacattgtggaacatctctgtgacattttgtgtgtgctgttgccgcatgttgcc	3816	

```
Search completed: June  2, 2002, 18:58:58
Job time: 11083 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:59:00 ; Search time 64.27 Seconds  
(without alignments)  
87.411 Million cell updates/sec

Title: US-09-663-600A-186  
Perfect score: 1174  
Sequence: 1 MASLGLVGYILGLLGLG.....PGQPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223.5	19.0	137	4	US-09-188-930-174
2	185.5	15.8	207	1	US-08-557-917A-2
3	185.5	15.8	207	4	US-09-084-153-2
4	185.5	15.8	207	4	US-09-084-079-2
5	179	15.2	218	4	US-09-084-079-5
6	91	7.8	160	3	US-08-600-430-5
7	87	7.4	160	3	US-08-600-430-6
8	87	7.4	161	1	US-08-557-917A-3
9	87	7.4	161	4	US-09-084-153-3
10	87	7.4	161	4	US-09-084-079-3
11	85	7.2	681	6	5194595-19
12	84.5	7.2	580	2	US-08-677-049-12
13	84	7.2	521	2	US-08-406-855A-19
14	84	7.2	521	3	US-09-206-899-19
15	83	7.1	144	1	US-08-518-474-8
16	82	7.0	157	2	US-08-883-070-3
17	82	7.0	468	2	US-08-390-000A-7
18	82	7.0	477	1	US-08-444-734A-4
19	82	7.0	477	1	US-08-087-772A-16
20	80.5	6.9	314	3	US-08-988-876-7
21	80.5	6.9	417	1	US-08-553-888A-3
22	79	6.7	158	3	US-09-010-809-19
23	78.5	6.7	277	1	US-08-118-270-68
24	78.5	6.7	277	5	PCT-US93-08528-68
25	78.5	6.7	381	4	US-08-876-798A-2
26	78	6.6	438	2	US-08-677-049-9
27	77.5	6.6	381	1	US-08-192-288-2

28 77.5 6.6 381 2 US-08-687-355A-2 Sequence 2, Appli  
29 77.5 6.6 381 4 US-09-200-673-16 Sequence 16, Appli  
30 77.5 6.6 160 2 US-08-883-070-7 Sequence 7, Appli  
31 76.5 6.5 475 1 US-07-686-591-4 Sequence 4, Appli  
32 76.5 6.5 475 1 US-07-970-715-4 Sequence 4, Appli  
33 75.5 6.4 227 2 US-08-254-493-1 Sequence 1, Appli  
34 75.5 6.4 227 2 US-08-253-751-6 Sequence 6, Appli  
35 75.5 6.4 227 2 US-08-453-925-6 Sequence 6, Appli  
36 75.5 6.4 227 4 US-08-403-253A-6 Sequence 6, Appli  
37 75.5 6.4 228 1 US-08-408-222B-1 Sequence 1, Appli  
38 75.5 6.4 360 4 US-09-045-583-51 Sequence 51, Appli  
39 75.5 6.4 501 1 US-08-722-001-14 Sequence 14, Appli  
40 75.5 6.4 501 2 US-08-467-568-9 Sequence 9, Appli  
41 75.5 6.4 501 2 US-09-030-582-9 Sequence 9, Appli  
42 75.5 6.4 572 5 PCT-US91-08177-11 Sequence 11, Appli  
43 75.5 6.4 647 1 US-08-218-943-1 Sequence 1, Appli  
44 74.5 6.3 330 3 US-08-115-753-1 Sequence 1, Appli  
45 74.5 6.3 419 3 US-08-115-753-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-188-930-174  
; Sequence 174, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Orust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.101c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Human  
US-09-188-930-174

Query Match 19.0%; Score 223.5; DB 4; Length 137;  
Best Local Similarity 35.7%; Pred. No. 4.2e-17;  
Matches 41; Conservative 30; Mismatches 43; Indels 1; Gaps 1;  
QY 67 YSTLLGLPADIQAAQAMVYTSSAISLACIISVVMRCITVFCQESRAKD-RVAVAGGVFF 125  
Db 7 YDSVLALSALQALTRALMVSVLGLFAMFVATPMKMKTRCGGDDKVKARIAMGGGIIF 66  
QY 126 ILGGLGFIPVAMNGLIRDFYSPVPSMKPEIGEALYGLISSLFSLIAGII 180  
Db 67 IVAGLAALVACSWYGHQIVTDYFNPLPTNPKYEFGEFPAIFGWSALVILGAL 121

RESULT 2  
US-08-557-917A-2  
; Sequence 2, Application US/08557917A  
; Patent No. 5756300  
; GENERAL INFORMATION:  
; APPLICANT: Bronstein, Jeff M.  
; APPLICANT: Seitz, Robert S.  
; APPLICANT: Lallone, Roger L.  
; TITLE OF INVENTION: Oligodendrocyte-specific Protein and Method for  
; TITLE OF INVENTION: Diagnosing and Treating Disease  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheldon & Mak

```

1 STREET: 225 S. Lake Avenue, 9th Floor
2 City: Pasadena
3 STATE: California
4 ZIP: 91101
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
8 COMPUTER: IBM compatible
9 OPERATING SYSTEM: Windows version 3.11
10 SOFTWARE: Wordperfect for Windows version 6.1
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/557_917A
14 FILING DATE: 14-NOVEMBER-1995
15 CLASSIFICATION: 436
16 ATTORNEY/AGENT INFORMATION:
17 NAME: Farah, David A.
18 REGISTRATION NUMBER: 38,134
19 REFERENCE/DOCKET NUMBER: 11201
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: (818)796-4000
22 TELEFAX: (818)795-6321
23 INFORMATION FOR SEQ ID NO: 2:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 207 amino acid residues
26 TYPE: amino acid
27 TOPOLOGY: linear
28 MOLECULE TYPE: protein
29
30 US-08-557-917A-2

```

```

Query Match          15.8%; Score 185.5; DB 1; Length 207;
Best Local Similarity 28.7%; Pred. No.1.2e-12;
Matches 64; Conservative 35; Mismatches 91; Indels 33; Gaps 10;

QY 1 MASIGLQVGYIILGLGLGLTLVAMLLPSW-KTSSVVGASIVTA-----VGFSGKGIWMEC 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MVACIAYGVFEVTSFVGMIGIIVTSTNDWVYTCGY---TIFPCRMDLGL-SKGIIMDC 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 55 ATHSGITOCQDIYSLFLGLPADIOAQMAMTSSAIISSLACTIISVGMRCITYCOE-SKA 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 VM-AGLGHCKRPLVDILLIPGYOACRALMAAAYLGLPAILLLTLVLPICIRMGHPGYA 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 114 KDRVAVAGVEFIELGLGLGFIIVANLNLGILRDFSPLYPSDMKPEI-----GEALYLGI 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 KYRRQQLAGVLLILLLALCAIYATIM-----FPVCAHHEITIVSFGYSLYAGW 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 169 ISSLESLIAG-IILCFSCSSQSRNRSN--YDAYYOQOPLATRSS 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 IGAVMCLVGGCYIVCCSGDAQSFGENRFFYSSGSSSPTHAKFA 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-084-153-2
: Sequence 2, Application US/09084153
: Patent No. 6147191
: GENERAL INFORMATION:
: APPLICANT: Bronstein, Jeff M.
: TITLE OF INVENTION: Oligodendrocyte-specific Protein
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheldon & Mak
: STREET: 225 S. Lake Avenue, 9th Floor
: CITY: Pasadena
: STATE: California
: ZIP: 91101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
: OPERATING SYSTEM: Windows 95
: SOFTWARE: WordPerfect for Windows version 8.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/084,153
: FILING DATE: 26-MAY-1998
: CLASSIFICATION:

```

```

? ATTORNEY/AGENT INFORMATION:
? NAME: Farah, David A.
? REGISTRATION NUMBER: 38,134
? REFERENCE/DOCKET NUMBER: 11201-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (626)796-4000
? TELEFAX: (626)795-6321
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 207 amino acid residues
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?-OS-084-153-2

```

```

Query Match 15.8% Score 185.5: DB 4; Length 207;
Best Local Similarity 28.7%: Pred. No. 1,2e-12;
Matches 64; Conservative 35; Mismatches 91; Indels 33; Gaps 10.

QY 1 MASLGLQVYLLGLGLGLVAMLLPSW-KTSSVYGAISYTA-----VGFSGKLMWEC 54
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MVATQLQVGFVTSFVGMIGIIVTSTINDMVYTCY---TIFCRKMDLGL-SKGLMADC 56

QY 55 ATHSGITQCDIYSYLLGLPADIQAAQAMMTSSAISLACIISVYGMCTYFCQE-SKA 113
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 57 VM-ATGLCHKRPLVDILLIGYQACRAMIAASVYGLPAILLLTLVPLCIRMGHPGYA 115

QY 114 KDRVAVAGVEPFIIGLLGFIIPANMLHGLRDFSPLPDMSKFEI-----GEALYLG 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 KYRRAGLQGLVLLILALCAIVATIM-----PPVCARHRIITIVSYGYSILYAGW 162

QY 169 ISSLFSLIAG-IILCFSCSSQNRNSN--YDAYQAQPLATRSS 208
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 163 IGAVMCLVGGCVIVCCSGDAOSFGENRFYSSGSSSPTFAKSA 205

RESULT 4
US-09-084-079-2
; Sequence 2, Application US/09084079
; Patent No. 6150136
; GENERAL INFORMATION:
; APPLICANT: Bronstein, Jeff M.
; APPLICANT: Seltz, Robert S.
; APPLICANT: Iallone, Roger L.
; TITLE OF INVENTION: Oligodendrocyte-specific Protein and Method for
; TITLE OF INVENTION: Diagnosing and Treating Disease
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 S. Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: California
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows version 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084, 079
; FILING DATE: 22-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farah, David A.
; REGISTRATION NUMBER: 38,134
; REFERENCE/DOCKET NUMBER: 11201-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626)796-4000
; TELEFAX: (626)795-6321
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acid residues

```

Query Match 15.2%; Score 179; DB 4; Length 218;  
Best Local Similarity 27.2%; Pred. No. 6.5e-12;  
Matches 62; Conservative 39; Mismatches 107; Indels

Qy	1	MASIGLQVQYILGLLGLGLTGLVAMLLPSSWKTSYVGSIVTAVGFSKGLHMECATHSG	60
Db	1	MLLLLSIIYLVHVAIVLL--FVSTISQW-----IVGNHATDLWNCSTSSSG	48
Qy	61	ITQCDIYSTLLGLPAD-IAQAQAMVTSSAIISSLACIIISVGMRCVVFQESRAKDRVAV	119
Db	49	-----NVHHCFFSSPNEWLQSVQATMILSIIFS-----ILSLFELFCOLEFTLTGKGRFYIT	99

Qy 120 AGGVFFLLGLGTFIPVA-----WNLHGILRDYSPPLVPDSMKFEIGALYGISS 171  
Db 100 --GIFOLLGLCYMSAAIYTVRHPHEWLNS-----DYSYGFAYILAWVAF 143  
Qy 172 LPSLIAGII 180  
Db 144 PLALLSGVI 152

## RESULT 7

US-08-600-430-6  
Sequence 6, Application US/08600430  
Patent No. 6124436  
GENERAL INFORMATION:  
APPLICANT: McClanahan, Terrill K.  
APPLICANT: Gorman, Daniel M.  
APPLICANT: Bolin, Laurel M.  
TITLE OF INVENTION: PURIFIED MAMMALIAN MONOCYTE ANTIGENS AND  
TITLE OF INVENTION: RELATED REAGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNA Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,430  
FILING DATE: 13-FEB-1996  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0517  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 160 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-600-430-6

Query Match 7.4%; Score 87; DB 3; Length 160;

Best Local Similarity 25.8%; Pred. No. 0.058; Matches 46; Conservative 26; Mismatches 70; Indels 36; Gaps 8;

Qy 12 ILGLGLGLTVLMLPSKTSYVGSIVTAVGFSKGLMMECATHTGOCDIYSTLL 71  
Db 2 LLLLLGLFLHIVLVLF-VSTIVSQMLV-GNGHTTDLMOCTTSALGAVQ-HCYSS-- 56  
Qy 72 GLPADIOAAQAMVMTSSAIISSVGMRCVPCQ-ESRAKDRVAVAGVFFIIGL 130  
Db 57 SVSEWLOSQVQATMILSVIFSVLALFL-----FFCOLFTLTKGGRYITGFFQIAGL 108  
Qy 131 LGTFIPVA-----WNLHGILRDYSPPLVPDSMKFEIGALYGISSLPSLIAGII 180  
Db 109 CVMSAAIYTVRHPHEWV-----NTDYSYGFAYILAWVAFPLALLSGII 152

## RESULT 8

US-08-557-917A-3  
Sequence 3, Application US/08557917A

Patent No. 5756300  
GENERAL INFORMATION:  
APPLICANT: Bronstein, Jeff M.  
APPLICANT: Seitz, Robert S.  
APPLICANT: Lallone, Roger L.  
TITLE OF INVENTION: Oligodendrocyte-specific Protein and Method for  
TITLE OF INVENTION: Diagnosing and Treating Disease  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheldon & Mak  
STREET: 225 S. Lake Avenue, 9th Floor  
CITY: Pasadena  
STATE: California  
ZIP: 91101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows version 3.11  
SOFTWARE: WordPerfect for Windows version 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,917A  
FILING DATE: 14-NOVEMBER-1995  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Farah, David A.  
REGISTRATION NUMBER: 38,134  
REFERENCE/DOCKET NUMBER: 11201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (818)796-4000  
TELEFAX: (818)795-6321  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 161 amino acid residues  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-557-917A-3

Query Match 7.4%; Score 87; DB 1; Length 161;

Best Local Similarity 25.8%; Pred. No. 0.058; Matches 46; Conservative 26; Mismatches 70; Indels 36; Gaps 8;

Qy 12 ILGLGLGLTVLMLPSKTSYVGSIVTAVGFSKGLMMECATHTGOCDIYSTLL 71  
Db 2 LLLLLGLFLHIVLVLF-VSTIVSQMLV-GNGHTTDLMOCTTSALGAVQ-HCYSS-- 56  
Qy 72 GLPADIOAAQAMVMTSSAIISSVGMRCVPCQ-ESRAKDRVAVAGVFFIIGL 130  
Db 57 SVSEWLOSQVQATMILSVIFSVLALFL-----FFCOLFTLTKGGRYITGFFQIAGL 108  
Qy 131 LGTFIPVA-----WNLHGILRDYSPPLVPDSMKFEIGALYGISSLPSLIAGII 180  
Db 109 CVMSAAIYTVRHPHEWV-----NTDYSYGFAYILAWVAFPLALLSGII 152

## RESULT 9

US-09-084-153-3  
Sequence 3, Application US/09084153  
Patent No. 6147191  
GENERAL INFORMATION:  
APPLICANT: Bronstein, Jeff M.  
TITLE OF INVENTION: Oligodendrocyte-specific Protein  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheldon & Mak  
STREET: 225 S. Lake Avenue, 9th Floor  
CITY: Pasadena  
STATE: California  
ZIP: 91101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM compatible

```

OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows version 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/084.153
FILING DATE: 26-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Farah, David A.
REGISTRATION NUMBER: 38, 134
REFERENCE/DOCKET NUMBER: 11201-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626)796-4000
TELEFAX: (626)795-6321
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 161 amino acid residues
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-153-3

```

Query Match 7.4%; Score 87; DB 4; Length 161;  
Best Local Similarity 25.8%; Pred. NO. 0.058;  
Matches 46; Conservative 26; Mismatches 70; Indels 36; Gaps 8;

```

QY 12 ILGLGLGLTIVAMLLPSWKTSSVVGASIVTAVFGSKGLMECATHTGTTQCDIYSTLL 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 LLLGLFLFLHIAVLVLF-VSTIVSQWL-VNGHTTDLWQNTCTSALGAVQ-HCYSS-- 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 GLPADIOQAAMMVTSSAISLLACITISVGMCRCTVFQ-ESRAKDRVAVAGGVFFLGLL 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 SVSEWLQVQATMILSVIFSLALF-----FFQLFTLTKGGRFVTGFFQILAGI 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 LGFTIPA-----WNLHGLRDFYSPYDPSMKKEIGEALYGLIISLSFLIAGII 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 CVMSAAAIYTVRHSNHV-----NTDYSYGAYILAWVAPLALLSGII 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10  
US-09-084-079-3  
; Sequence 3, Application US/09084079

Patent NO. 8130136  
 GENERAL INFORMATION:  
 APPLICANT: Bronstein, Jeff M.  
 APPLICANT: Seitz, Robert S.  
 APPLICANT: Lallone, Roger L.  
 TITLE OF INVENTION: Oligodendrocyte-specific Protein and Method for  
 TITLE OF INVENTION: Diagnosing and Treating Disease  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sheldon & Mak  
 STREET: 225 S. Lake Avenue, 9th Floor  
 CITY: Pasadena  
 STATE: California  
 ZIP: 91101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows version 8.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/084,079  
 FILING DATE: 22-MAY-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Farah, David A.  
 REGISTRATION NUMBER: 38,134  
 REFERENCE/DOCKET NUMBER: 11201-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (626)796-4000  
 TELEFAX: (626)795-6321  
 INFORMATION FOR SEQ ID NO: 3:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acid residues
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-084-079-3

```

Query Match	7.4%	Score 87;	DB 4;	Length 161;
Best Local Similarity	25.8%;	Pred. NO. 0.058;		
Matches 46;	Conservative 26;	Mismatches 70;	Indels 36;	Gaps 8;

[illegible]

```

RESULT 11
5194595-19
;Patient No. 5194595-19
;APPLICANT: WAT
;TITLE OF INVENT
;IMMUNOGENIC SEGMENT
;SYNCTIL VIRUS
;NUMBER OF SEQU
;CURRENT APPLIC
;APPLICATION
;FILING DATE:
;PRIOR APPLICAT
;APPLICATION
;FILING DATE:
;SEQ ID NO:19
;LENGTH: 681
5194595-19

```

Query Match 7.2%; Score 85; DB 6; Length 681;  
Best Local Similarity 20.1%; Pred. NO. 0.77;  
Matches 59; Conservative 48; Mismatches 92; Indels 94; Gaps 17;

	Qy	1	MASIGQLVGYI--LGLLLGTLVLAMLLPSWK-----TSSYVGASIVTVAGFSKGLW	51
	Dd		: : : : :   : : :   : :   : : :   :	
	Dd	289	MSIIKEEVLAVVQLPLYGVIDT-----PCWLHSTPLCTTTKEGSNICLTRTRDGRWY	342
	Qy	52	ME-----CATHSTGTQCDDYSTLLGPLADLQAQM-	84
	Dd	343	CDNAGSVSFFPQAEKTKVQSNRVF-CUTMNSLT-LFSEINLCNWDIFNPKYDCKIMTSKT	400
	Qy	85	-VTSSAITSACIISVVG-MRCTVF-----CO--ESRAKRADVAVAGGVFFT-	126
	Dd	401	DVSSTSVITSLGAIVSCYCGKTKCFASNKNRGIIKTFSGCDYVSKNGMDTVSVGNTLYYNV	460
	Qy	127	-LGGLLGFIL-PVANLHGILRDFYSPLVDPMKMFEALVGLISS-	171
	Dd	461	KQEGKSILYVKGEPII-----NFYDLVPFDSEFD-----QLGPSNPSEITSQITT	507
	Qy	172	-LFSLIAGIILCFSCSQNRNSNYDAYAQAPLATRSSRPQGPP-KVKSEFN	222
	Dd	508	I LASTPGVKASTLQSSTVKTNK-TTTQTQPSPKITQRONKPSPKNNDH	558

RESULT 12  
US-08-677-049-12  
; Sequence 12, Application US/08677049  
; Patent No. 5858707

```

GENERAL INFORMATION:
APPLICANT: Guimaraes, M. Jorge
APPLICANT: Bazan, J. Fernando
APPLICANT: McClanahan, Terrill K.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNA Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/677,049
FILING DATE: 03-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000,788
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 207..236
OTHER INFORMATION: /note="Encompasses TM 4 of Figure
OTHER INFORMATION: 4"
NAME/KEY: Region
LOCATION: 437..499
OTHER INFORMATION: /note="Encompasses TM 9 and TM 10
OTHER INFORMATION: of Figure 4"
US-08-677-049-12

Query Match 7.2%; Score 84.5; DB 2; Length 580;
Best Local Similarity 25.2%; Pred. No. 0.7;
Matches 59; Conservative 29; Mismatches 93; Indels 53; Gaps 11;

OY 8 LVGVIIGLGLGLTIVAMLPMSKTSYVAGSIYTVANGFSKGLMEAT-----HSGI 61
DB 110 LVSTSLVSLGLSVAQFRLHYKTRVYTGIVSVGTSGT-ITVATGFNOMYSTGY 168
OY 62 TOCDITSTLLGLPADIOAQAOMVWTSATSLACIIISVGMRCVTFQESRAKDRVAAG 121
DB 169 CPVDGSGNRLPCP---KGYQALLATSLCLSLLEIGLSFMSRL-----KALPPITYG 219
OY 122 GVFFILG-GLGFIPIVAM-----NLHGL---RDFYSPVLPDSMKF-EIGEALYL 166
DB 220 PVFVLIGASLIGNAMKMGSGTCSNPGGALCPADAPHPLPMSGAETIGIGFLVFA 279
OY 167 GII-----SSLSLAGIILCFSCSQRNRSTYD--AYQAOPLAT 205
DB 280 TILCFRFGSPIMKSCAVIVGLVGCIVAAAC-----GYFDRSGIDAPVAVS 326
```

```

RESULT 13
US-08-406-855A-19
Sequence 19, Application US/08406855A
Patent No. 5861309
GENERAL INFORMATION:
APPLICANT: Jonathan A. Bard et al.
TITLE OF INVENTION: DNA Encoding Human Alpha 1 Adrennergic
Receptors and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406,855A
FILING DATE: 21-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41337-A-PCT-US/JPW/KDB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-406-855A-19

Query Match 7.2%; Score 84; DB 2; Length 521;
Best Local Similarity 20.5%; Pred. No. 0.68;
Matches 58; Conservative 36; Mismatches 79; Indels 110; Gaps 14;

OY 18 LLGTVVAMLLPMSKTSY-----VGASIVTVANGFSKGLM--- 52
DB 5 LRSVMAGYLSMRTPYRSTEMVQRLRMEAYOHSSTAANGVGLVVSAGVGCVFLAIF 64
OY 53 -----ECATHSTGITO-----CDIYSTLLGLP----- 74
DB 65 ILTAVAGNLLVLSVACNHNLOTVNYEIVNLAVADLLLSAIVLPFSATMEVIGFWAFGR 124
OY 75 --ADIOQAOMVWTSATSLACIIISV---VGMRCV---EQESRADRVAVAGVFF 125
DB 125 TFCDDVAAVADVLCYASLISL-CTISVDRIYGVGRHSLSKYPATMTERKAATIALLMAVAL 183
OY 126 I--LGLGLGFIPIVANNLIGILDRFSPLVPD---SMKEIGEGALYLGIISSFLIAGI 179
DB 164 VVSVPPLG-----WK-----EPVPPDERFCGITEVEGYAIFSSCS--FYLPMAV 227
OY 180 ILCFSCSSQRNRSTYDAYQAOPLATRS-----SPRPGQPPRY 217
DB 228 IVVMTCR-----YVVARSTTRSLDEAGIKRREPGRASEV 260

RESULT 14
US-09-206-899-19
Sequence 19, Application US/09206899
Patent No. 6083705
```

GENERAL INFORMATION:  
; APPLICANT: Jonathan A. Bard et al.  
; TITLE OF INVENTION: DNA Encoding Human Alpha 1 Adrenergic  
; TITLE OF INVENTION: Receptors and Uses Thereof  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/206,899  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/406,855  
; FILING DATE: 21-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41337-A-PCT-US/JPW/KDB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-206-899-19

Query Match 7.28; Score 84; DB 3; Length 521;  
Best Local Similarity 20.58; Pred. No. 0.68;  
Matches 58; Conservative 36; Mismatches 110; Gaps 14;  
QY 18 LLGLTLVAMLLPSWKTSSY-----VGASIVTAVGFSKGLWM--- 52  
DB 5 LRSVMAGYSEWTPYRSTEMYORLMEAVQHSSTAAVGLVVSAGVGVGVLAA 64  
QY 53 -----ECATHSTGITQ-----CDIYSTLLGLP----- 74  
DB 65 ILTAVAGNLLVILSVACNRHLQTVTNFYFVNLAVADLLLSAVALPFSATMEVLGFWAFGR 124  
QY 75 --ADIQAAQAMVTSIAISSLACIISV---VGMRCVTV-----FCQESRAKDRVAVAGGVFF 125  
DB 125 TFCVDVAAADVLCCTASILSL-CTISVDYVGVVRHSLKYPAIMTERKAAAILALLMAVAL 183  
QY 126 I--LGGLLGFTPVANLHGLRDFYSLVPD-----SMKFEIGEALYGLIISLFLIAGI 179  
DB 184 VVSVGPLLG-----WK-----EVPDPDFCGITEEVGYAIFSSVCS--FYLPMVAV 227  
QY 180 ILCFSCSQNRNRSNYDAYQAQPLATRS-----SPRPGQPPKV 217  
DB 228 IVVMYCR-----VYVVARSTTRSLKAGIKRPGKASEV 260

RESULT 15  
US-08-518-474-8  
; Sequence 8, Application US/08518474  
; Patent No. 5599920  
; GENERAL INFORMATION:  
; APPLICANT: Patel, Pragna I.  
; APPLICANT: Suter, Ueli

APPLICANT: Snipes, G. J.  
APPLICANT: Welcher, Andrew  
APPLICANT: DeLeon, Marino  
APPLICANT: Lupski, James R.  
APPLICANT: Shooter, Eric M.  
TITLE OF INVENTION: PERIPHERAL MYELIN PROTEIN CODING  
TITLE OF INVENTION: SEQUENCE AND METHOD  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,474  
FILING DATE: 21-AUG-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,488  
FILING DATE: 19-SEP-1994  
APPLICATION NUMBER: US 07/879,623  
FILING DATE: 06-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul, Thomas D.  
REGISTRATION NUMBER: 32,714  
REFERENCE/DOCKET NUMBER: D-5331-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713/651-5151  
TELEFAX: 713/651-5246  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-518-474-8

Query Match 7.18; Score 83; DB 1; Length 144;  
Best Local Similarity 29.9%; Pred. No. 0.14;  
Matches 35; Conservative 17; Mismatches 47; Indels 18; Gaps 6;  
QY 12 ILGLLGLLTLVAMLLPSWKTSSYVVGASIVTAVGFSKGLWMECATHTSGITQCDIYSTLL 71  
DB 2 LLLLGLILFHIAVLVLLF-VSTIVSQWLV-GNGHTTDLWQNTTSALGAVQ-HCYSS-- 56  
QY 72 GLPADIQAAQAMVTSIAISSLACIISVVGMRCTVFCQESRAKDRVAVAGGVFFILG 128  
DB 57 SVSEWLSQVQATWILSVIFSLALFL-----FFCQLF-----TLTKGGRFYING 100

Search completed: June 2, 2002, 20:13:56  
Job time: 4496 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:58:20 ; Search time 93.15 seconds  
(without alignments)  
237.258 Million cell updates/sec

Title: US-09-663-600A-186  
Perfect score: 1174  
Sequence: 1 MASLGLQVGYILGLLGLG.....PGOPPKVKSEFNSYLTGYV 230  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	423.5	36.1	280	2 A39484	androgen-withdrawal
2	100	8.5	400	2 D97643	probable efflux pr
3	100	8.5	400	2 AG2866	MFS permease (drug
4	99.5	8.5	330	2 C69648	2-keto-3-deoxygluc
5	99.5	8.5	463	2 F90285	metabolite transpo
6	97.5	8.3	484	2 C75609	amino acid ABC tra
7	95.5	8.1	458	2 H71657	NADH dehydrogenase
8	93	7.9	523	2 T11916	NADH dehydrogenase
9	91	7.8	160	2 JN0503	peripheral myelin
10	90.5	7.7	487	2 T19237	hypothetical prote
11	90.5	7.7	617	2 F71019	hypothetical prote
12	88.5	7.5	429	2 A97241	permease [imported
13	88.5	7.5	521	2 A95349	amino acid permeas
14	88.5	7.5	593	2 H71283	conserved hypothet
15	88	7.5	160	2 A41144	growth arrest-rela
16	88	7.5	421	2 A99309	membrane transport
17	88	7.5	906	2 G90281	conserved hypothet
18	87.5	7.5	268	2 A71086	hypothetical prote
19	87.5	7.5	420	2 H96534	probable Na/H+ an
20	87.5	7.5	519	2 T15364	hypothetical prote
21	87.5	7.5	690	2 S35251	probable membrane
22	87	7.4	157	2 G02355	tumor-associated m
23	87	7.4	160	2 S21721	growth arrest-spec
24	87	7.4	223	2 E87654	hypothetical prote
25	87	7.4	318	2 AG3576	iron(III) dicitrat
26	87	7.4	414	2 F65097	hypothetical prote
27	87	7.4	414	2 C91125	probable transport
28	87	7.4	414	2 B85970	probable transport
29	86.5	7.4	255	2 E64604	conserved hypothet

30	86.5	7.4	402	2 A11853	sodium-dependent n
31	86.5	7.4	432	2 A70714	hypothetical prote
32	86.5	7.4	458	2 D97799	NADH dehydrogenase
33	86.5	7.4	497	2 AD3525	amino acid permeas
34	86.5	7.4	547	2 B64963	membrane protein y
35	86	7.3	398	2 J00113	2-keto-3-deoxygluc
36	86	7.3	409	2 S29124	membrane glycoprot
37	86	7.3	562	2 T52481	cytochrome-c oxida
38	85.5	7.3	216	2 I38474	olfactory receptor
39	85.5	7.3	371	2 D75266	cell division prot
40	85.5	7.3	414	2 AG0894	probable membrane
41	85.5	7.3	445	2 T11097	NADH dehydrogenase
42	85	7.2	233	2 T31865	hypothetical prote
43	85	7.2	502	2 B70845	probable sygar tra
44	85	7.2	531	2 T11074	NADH dehydrogenase
45	84.5	7.2	454	2 C86766	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

A39484  
androgen-withdrawal apoptosis protein RVPL, prostatic - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 04-Mar-2000  
C:Accession: A39484  
R:Briehl, M.M.; Miesfeld, R.L.  
Mol. Endocrinol. 5, 1381-1388, 1991  
A:Title: Isolation and characterization of transcripts induced by androgen withdrawal  
A:Reference number: A39484; MUID:92130987  
A:Accession: A39484  
A:Molecule type: mRNA  
A:Residues: 1-280 <BRI>  
A:Cross-references: GB:M74067; NID:g205857; PIDN:AAA41760.1; PID:g205858  
C:Genetics:  
A:Gene: RVP.1  
C:Superfamily: rat androgen-withdrawal apoptosis protein RVPL

Query Match 36.1%; Score 423.5; DB 2; Length 280;  
Best Local Similarity 37.7%; Pred. No. 4.2e-29;  
Matches 84; Conservative 50; Mismatches 74; Indels 15; Gaps 3;

OY	3	SLGLQLVGYILGLLGLTLVAMLPSSVVGASIVTAVGFSGLMMECATSTGIT	62
Db	2	SMSLEITGTSVLAVGLWELTIVCCALPMWRVSAFIGSIIITQITWEGLMWNC-VQSTGOM	60
OY	63	OCDIYSTLLGLPADIOAAQAMMYTSSAISLACIIISVGMRICTVFCQESRAKDRVAVGG	122
Db	61	OCKMYDSLIALPODLQAARALIVVSTLLAAFGLLVALVGAOCTNCVODETAKAKITIVAG	120
OY	123	VFFILGGLGFTPVANWHLGILRDFYSPVDPMSKKEIGALYGIISLSFSLIAGIILC	182
Db	121	VLFLLAAVLTLPVSVSANTIIIRDFYNPLVPEAKREMGTLGVGWAALQLLGGALLC	180
OY	183	FSCSQNRNSNYDYAOAOPLAT-----RSSPRPG	212
Db	181	CSCPPREKYAPTAKILYSA-PRSTGPGTGTAYDRKTTSERPG	222

##### RESULT 2

D97643  
probable efflux protein cjl687 [imported] - Agrobacterium tumefaciens (strain C58, Ce  
C:Species: Agrobacterium tumefaciens  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: D97643  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouellet, B.; Goldm  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194  
A:Accession: D97643





```

Oy 60 ----GTTGCDIYSTLLGLPADIQAAQAMMTSSAIISSLACIIISVGMRCYVFCQESRAKD 115
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 QFTDGVCSPPWVYS-----DVTPLAKA--ASSSSFSALNPAIADGI-----120

Oy 116 RAVAGVFFILGLGCFIP-----VAM-NLHGILRDYSPPLVDSMKFEIGALYLGIISS 170
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 -----SFPFIILTTL-LVPICILVSNMNIETVYKEY-----CIAFLVETLMTLVFS 166

Oy 171 SLESLI 176
      1 :
Db 167 VLDLL 172

RESULT 9
JN0503
Peripheral myelin protein 22 - human
N.Alternate names: Charcot-Marie-Tooth; GAS-3 protein; growth arrest-specific protein 3;
C.Species: Homo sapiens (man)
C.Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Jun-2000
C.Accession: JN0503; JCI190; A56697; S25537
R.Edoml, P.; Martinotli, A.; Colombo, M.P.; Schneider, C.
Gene 126, 289-290, 1993
A.Title: Sequence of human GAS3/PMP22 full-length cDNA.
A.Reference number: JN0503; M01D:93246261
A.Accession: JN0503
A.Molecule type: mRNA
A.Residues: 1-160 <ED0>
A.Cross-references: GB:DL1428; NID:q182984; PIDN:AAA58495.1; PID:q182985
R.Hayasaka, K.; Himoro, M.; Nanao, K.; Sato, W.; Mura, M.; Uyemura, K.; Takahashi, E.;
Biochem. Biophys. Res. Commun. 186, 827-831, 1992
A.Title: Isolation and sequence determination of cDNA encoding PMP-22 (PAs-II/SRI3/GAS-3
A.Reference number: JCI190; M01D:92360032
A.Accession: JCI190
A.Molecule type: mRNA
A.Residues: 1-160 <YAL>
A.Cross-references: GB:DL1428; NID:q220009; PIDN:BA0195.1; PID:q220010
R.Valentijn, L.J.; Baas, F.; Wolterman, R.A.; Hoogendijk, J.E.; van den Bosch, N.H.A.; Z
Nature Genet. 2, 288-291, 1992
A.Title: Identical point mutations of PMP-22 in Trembler-J mouse and Charcot-Marie-Tooth
A.Reference number: A56697; M01D:93265161
A.Accession: A56697
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-160 <YAL>
A.Cross-references: GB:S61788
R.Colombo, M.
submitted to the EMBL Data Library, May 1992
A.Reference number: S25537
A.Accession: S25537
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 61-160 <COL>
A.Cross-references: EMBL:X65968; NID:q31652; PIDN:CAA46781.1; PID:q31653
C.Genetics:
A.Gene: GDB:PMP22; GAS3
A.Cross-references: GDB:134190
A.Map position: 17p12-17p11.2
C.Superfamily: growth arrest-specific protein
C.Keywords: myelin; transmembrane protein

Query Match 7.8%; Score 91; DB 2; Length 160;
Best local Similarity 21.7%; Pred. No. 0.96; Mismatches 69; Indels 46; Gaps 8;
Matches 41; Conservative 33;

Oy 1 MASIGIOLVYIIGLLGLTIVAMLLPSWKTSSYGVASTYAVGFSKGLMMECATHTSTG 60
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLLLLSIIVLHVAIVL--FVSTVSGW-----IVGNGHNDLMQNKSTSSSG 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 61 TTAGCIVSTLLGLPAD-IQAQAAMMTSSAIISSLACIIISVGMRCYVFCQESRAKDRVAV 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 49 ----NVHICFSSPNEMLQSVQATMILSTIIS-----ILSLFFPCQLPTITGGRFYIT- 99
Oy 120 AGVFFILGLGFIPIVA-----WNLHGILRDYSPPLVDSMKFEIGALYLGIISS 171
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 --GIPOILAGLCVMSAAIYTVRHPDEMLNS-----DYSGFAYILLAMVAF 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy 172 LPSLIGIT 180
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 PLALLSGVI 152

RESULT 10
T19237
hypothetical protein C27A7.6 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C.Accession: T19237; T19502
R.Harris, B.
submitted to the EMBL Data Library, March 1997
A.Reference number: Z19095
A.Accession: T19237
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-487 <WIL>
A.Cross-references: EMBL:Z92825; PIDN:CAB07317.1; GSPDB:GN00023; CESP:C27A7.6
R.Harris, B.
submitted to the EMBL Data Library, October 1996
A.Reference number: Z19132
A.Accession: T19502
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-487 <W12>
A.Cross-references: EMBL:Z81041; PIDN:CAB02792.1; GSPDB:GN00023; CESP:C27A7.6
A.Experimental source: clone C27A7
C.Genetics:
A.Gene: CESP:C27A7.6
A.Map position: 5
A.Introns: 30/3; 51/3; 87/3; 109/3; 183/1; 215/3; 264/2; 302/2; 330/3; 349/3; 372/1;

Query Match 7.7%; Score 90.5; DB 2; Length 487;
Best local Similarity 28.7%; Pred. No. 3.3;
Matches 51; Conservative 17; Mismatches 79; Indels 31; Gaps 7;

Oy 5 GLOIVGIIIGLLGLTIVAMLLP--SWKTSSYGVASTYAVGFSKGLMMECATHTST 59
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 GTEMAGYPTAVCAIVGTLTSLAGHIDRTKRFETIRVCTVGFSGSVITLRPFLNQPT 356
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy 60 GTTQCDIYSTLLGLPADIQAAQAMMTSSAIISSLACIIISV-VGMRCYVFCQESRAKDRVA 118
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 GLFDSIIYVTICG-----CLGAFSIRPFIQVEIGVETTFYVME-----A 396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy 119 VAGVFFILGLGIF-IPVAMNLHGILRDYSPPLVDSMKFEIGALYLGIISSLSFL 175
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 TSSGVLVIFGSLFMFLIPFAQNYTEKLHLFYA-----QSWKFAIDVYCGSIYSVILSL 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
F71019
hypothetical protein PH1451 - Pyrococcus horikoshii
C.Species: Pyrococcus horikoshii
C.Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C.Accession: F71019
R.Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohitoku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A.Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A.Reference number: A71000; M01D:98344137
A.Accession: F71019
A.Status: preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-617 <KAN>

```

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30558.1; PID:g3257875  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1451  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

Query Match 7.7%; Score 90.5; DB 2; Length 617;  
Best Local Similarity 24.4%; Pred. No. 4.2;  
Matches 52; Conservative 33; Mismatches 81; Indels 47; Gaps 9;

QY 5 GLQVGVILGLGL-----GTLVAMLLPSWKTSY-----VCASVTAVGFS-- 47  
DB 273 GTPFGYIIAFLGLGLTIIGGVLAALQEDIRKLFAYSSISQVGVILVGLGIGTSLGIEAA 332

QY 48 -----KGL-WMECAT--HSTGITQCDIYSTLLGLPADIAQAAMVTSS-----AI 90  
DB 333 IYHAISHALFKGLFFLVAVIYTGTEKDFGGL-----AEKMPFFAMAFIAI 383

QY 91 SSLACIISVVMRCCTVFCQSRADRVAVAGGVFFILGGLGFIPVAMNHLGILRDFYSP 150  
DB 384 LSLAGPPLVGFASKLIFEAVISOKLPILGMLF-FCGALGFYVLIIRFTVAVWFGQRP 442

QY 151 LVPDSMKFEIGEALYGLIISL---FSLIAGII 180  
DB 443 DIETDKDAPLPLAIGMIGLTLNVVGFAPGLV 475

RESULT 12  
A97241  
permease [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: A97241  
R.Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: A97241  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-429 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK80716.1; PID:ql5025810; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2772  
C:Superfamily: conserved hypothetical protein H0125

Query Match 7.5%; Score 88.5; DB 2; Length 429;  
Best Local Similarity 22.8%; Pred. No. 4.3;  
Matches 41; Conservative 38; Mismatches 76; Indels 25; Gaps 8;

QY 13 LGLLGLG-TLVAMLLPSWKTSYSGASIVTAVGF-----SKGLWMECAT-----S 58  
DB 79 MGLNALFTYICIQLFHSWKA--LAASLIEGIIFLVNVFKIRQLILDSVPQTLKYAIS 136

QY 59 TGTQCDIYSTLLGL--PADIAQAAMVTSSAISLACIISVVMRCCTVFCQSRADRV 116  
DB 137 IGI---GFFITFGLODAGIVSGKPLVTLASLSPAVLAVIGVLLIYLVNKNIKGS 193

QY 117 VAVAGGVFFILGGLGFIPVAMNHLGILRDFYSPVLPDSMKFEIGEALYGLIISLFSLI 176  
DB 194 FVIGMLVIYVLGIFG---VAKAPSGIV-SFPSPVAPVFLQDFKSAVIGIVPVTML 249

RESULT 13  
A99549  
amino acid permease [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C:Accession: A99549  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: A99549  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <KUR>  
A:Cross-references: GB:AL445566; PID:gl4089711; PIDN:CAC13470.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPV\_2970  
A:Genetic code: SGC3

Query Match 7.5%; Score 88.5; DB 2; Length 521;  
Best Local Similarity 21.2%; Pred. No. 5.3;  
Matches 50; Conservative 38; Mismatches 63; Indels 85; Gaps 11;

QY 1 MASLGLQVLGVYILGLLGLTLVAMLL-----PSWKTSSYVGA--SIVTA 43  
DB 1 MSEKTKKVGFFVALSMLGVSVGIGIFPKNGSVGRAVEHNGTSLLANVFGGIIISLAA 60

QY 44 VGFSK-----GLWMECATHSYG-----ITQCDIYSTLLG----- 72  
DB 61 INFSEISFLNKTAKIAGIGNW---SHKVGDKRFGYFVSISFTLYSGIIILGFFTAEI 116

QY 73 -----LPADIAQAQ-AMMVTSSAISLACIISVVMRCCTVFCQSRADRVAVAGGV 124  
DB 117 PFHMLNLGANIKMEMHSLVLGTVISFIILNIVSIK-----ASGVF 160

QY 125 FILGGLGFIP-VAMNHLGI-----LRDFYSPVLPDSMKFEIGEALYGLIISL 172  
DB 161 QVITILAFPLIATILVGLIFVSTHTLEDGANAFVFNKTK-KIGEFSEANVIAAL 215

RESULT 14  
H71283  
conserved hypothetical integral membrane protein TP0771 - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
C:Accession: H71283  
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: H71283  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-593 <COL>  
A:Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65739.1; PID:g332  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0771

Query Match 7.5%; Score 88.5; DB 2; Length 593;  
Best Local Similarity 23.2%; Pred. No. 6;  
Matches 45; Conservative 36; Mismatches 72; Indels 41; Gaps 8;

QY 4 LGLQVLGVYILGL----LGLLGLTLVAMLLPSWKTSYV-----GASIVTAVGSKGLWMEC 54  
DB 177 LGESTMGF--GULFTGCLGLLSLIPPL--SVEELSPLKIAVEDRALSVFVLGSGFLVAV 232

QY 55 ATHSTGITQCDIYSTLLGLPADIAQAAMVTSSAISLACIISVVMRCCTVFCQSRADRV 114  
DB 233 ILHSSATATVILTMFAGVGIVGEFAAASVLGSGNCGSTIDAAIAAIG-----SKLN 283

QY 115 DRVAVAGGVFFILGGLGFIPVAMNHLGILRDFYSP-----LVPDSMKFEIGEALYGL 167  
DB 115 DRVAVAGGVFFILGGLGFIPVAMNHLGILRDFYSP-----LVPDSMKFEIGEALYGL 167

DB 284 ARRAAVHVLNFVGLVFLM-----PFHPVALLCVITPRKNSGPD-NITVRLA 331

OY 168 IISLFLIAGIIL 181

DB 332 LFHSMENIVNTIIV 345

RESULT 15

A41144

growth arrest-related myelin protein precursor, sciatic nerve - rat

N:Alternate names: Schwann cell membrane glycoprotein, SAG

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Jun-1992 #sequence, revision 30-Jun-1992 #text, change 29-Sep-1999

C:Accession: A41144; S18550; A44826

R:Welcher, A.A.; Suter, U.; De Leon, M.; Snipes, G.J.; Shooter, E.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 7195-7199, 1991

A:Title: A myelin protein is encoded by the homologue of a growth arrest-specific gene.

A:Reference number: A41144; MUID:91334432

A:Accession: A41144

A:Molecule type: mRNA

A:Residues: 1-160 <WEL>

A:Cross-references: GB:M69139; NID:9207063; PIDN:AA73063.1; PID:9207064

R:Spreyer, P.; Kuhn, G.; Hanemann, C.O.; Gillen, C.; Schaal, H.; Kuhn, R.; Lemke, G.; M

EMBO J. 10, 3661-3668, 1991

A:Title: Axon-regulated expression of a Schwann cell transcript that is homologous to a

A:Reference number: S18550; MUID:92037527

A:Accession: S18550

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-29, 'E', '31-160 <SPR>

A:Cross-references: EMBL:X62431; NID:955903; PIDN:CAA44297.1; PID:955904

R:Diaperink, M.E.; O'Neill, A.; Magnoni, G.; Wollmann, R.L.; Heinrichson, R.L.; Zucker-N

J. Neurosci. 12, 2177-2185, 1992

A:Title: SAG: a Schwann cell membrane glycoprotein.

A:Reference number: A44826; MUID:92300442

A:Accession: A44826

A>Status: preliminary

A:Molecule type: protein

A:Residues: 1-21, 'K', '23-25, 'X', '27, 'R', '29-30 <DIE>

A:Note: sequence extracted from NCBI backbone (NCBIP:106581)

C:Comment: The predicted signal sequence may not be cleaved in some cell types.

C:Superfamily: growth arrest-specific protein

C:Keywords: glycoprotein; Schwann cell; transmembrane protein

Query Match 7.5%; Score 88; DB 2; Length 160;

Best Local Similarity 25.8%; Pred. No. 1.7;

Matches 46; Conservative 26; Mismatches 70; Indels 36; Gaps 8;

OY 12 ILGLGLCTLVAMLPMTSSVGSIVTAVGFSKGLMMECATSTGTQCDIYSTLL 71

DB 2 LLLGLGLFLHIAVLVLF-VSTIVSQMLV-GNGHRTDLMQNTTSALGAVQ-HCYSS-- 56

OY 72 GLPADIQAAQAMWVTSSAISLACIISVGMKCTVFCQ-ESRAKDRVAVAGVFILGL 130

DB 57 SVSEMLQSVQATFMTLSVFLSLFL-----FCQLFTLTKGGRFYITGVFOILL 108

OY 131 LGFIPVA-----WNHGLTLDFTSPPLVDSMKFETGALVIGIISLFLAGII 180

DB 109 CVMSAAAIYTVRHSSEWVNN-----DYSYGFAVITLAWVAFLALSLGII 152

Search completed: June 2, 2002, 20:12:43  
Job time: 4463 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 20:09:12 ; Search time 65.66 Seconds  
(without alignments)

135.630 Million cell updates/sec

Title: US-09-663-600A-186

Perfect score: 1174

Sequence: 1 MASLGLVLGILGLLGLG.....PCQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1174	100.0	230	1 CLD2_HUMAN	P57739 homo sapien
2	1119	95.3	230	1 CLD2_CANFA	Q95km6 canis famil
3	1096	93.4	230	1 CLD2_MOUSE	Q88552 mus musculu
4	571	48.6	239	1 CLDE_HUMAN	Q95500 homo sapien
5	561.5	47.8	239	1 CLDE_MOUSE	Q920s3 mus musculu
6	459.5	39.1	208	1 CLDY_BRARE	Q9yh9l brachydanio
7	450.5	38.4	218	1 CLD3_CANFA	Q95km5 canis famil
8	450	38.3	220	1 CLD3_HUMAN	Q15551 homo sapien
9	446	38.0	210	1 CLD4_MOUSE	Q35034 mus musculu
10	443.5	37.8	209	1 CLD4_CERAE	Q19005 cercopithec
11	443	37.7	219	1 CLD6_MOUSE	Q92262 mus musculu
12	442.5	37.7	209	1 CLD4_HUMAN	Q14493 homo sapien
13	441	37.6	217	1 CLD9_HUMAN	Q95484 homo sapien
14	441	37.6	219	1 CLD3_MOUSE	Q920g9 mus musculu
15	437	37.2	219	1 CLD3_RAT	Q63400 rattus norv
16	437	37.2	220	1 CLD6_HUMAN	P56747 homo sapien
17	434	37.0	217	1 CLD9_MOUSE	Q920s7 mus musculu
18	431.5	36.8	211	1 CLD7_MOUSE	Q92261 mus musculu
19	427.5	36.4	211	1 CLD7_HUMAN	Q95471 homo sapien
20	407.5	34.7	211	1 CLD1_HUMAN	Q95832 homo sapien
21	401.5	34.2	211	1 CLD1_MOUSE	Q88551 mus musculu
22	400	34.1	209	1 CLD2_BRARE	Q9yh90 brachydanio
23	399.5	34.0	215	1 CLDX_BRARE	Q9yh92 brachydanio
24	391.5	33.3	219	1 CLDI_HUMAN	P56880 homo sapien
25	387.5	33.0	211	1 CLDI_MOUSE	P56745 rattus norv
26	386.5	32.9	225	1 CLD8_MOUSE	Q92260 mus musculu
27	378	32.2	218	1 CLD5_HUMAN	Q00501 homo sapien
28	376	32.0	224	1 CLDH_HUMAN	P56750 homo sapien
29	375.5	32.0	191	1 CLD7_RAT	Q92111 rattus norv
30	372.5	31.7	225	1 CLD8_HUMAN	P56748 homo sapien
31	367	31.3	218	1 CLD5_MOUSE	Q54942 mus musculu
32	353	30.1	206	1 CLD5_RAT	Q9jkd6 rattus norv
33	327.5	27.9	228	1 CLD5_HUMAN	P56746 homo sapien

34	291	24.8	228	1 CLDA_HUMAN	P78369 homo sapien
35	290	24.7	231	1 CLDA_MOUSE	Q920s6 mus musculu
36	274	23.3	264	1 CLDI_MOUSE	P56857 mus musculu
37	264.5	22.5	261	1 CLDI_HUMAN	P56856 homo sapien
38	223.5	19.0	211	1 CLDD_MOUSE	Q920s4 mus musculu
39	192.5	16.4	207	1 CLDB_HUMAN	Q75508 homo sapien
40	185.5	15.8	207	1 CLDB_MOUSE	Q60771 mus musculu
41	154.5	13.2	122	1 CLDF_MOUSE	Q920s5 mus musculu
42	153.5	13.1	254	1 CLDG_BOVIN	Q9xt98 bos taurus
43	149.5	12.7	305	1 CLDG_HUMAN	Q9y517 homo sapien
44	136.5	11.6	244	1 CLDC_HUMAN	P56749 homo sapien
45	101.5	8.6	323	1 CCG2_HUMAN	Q9y698 homo sapien

#### ALIGNMENTS

RESULT 1

ID	CLD2_HUMAN	STANDARD;	PRT;	230 AA.
AC	P57739;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Claudin-2.			
GN	CLDN2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Epithelium;			
RA	Reinecker H.-C., Sakaguchi T., Golden H.M.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,			
RA	Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,			
RA	Yu J., Han L.H.;			
RT	*Novel human cDNA clone with function of inhibiting cancer cell			
RT	growth.*;			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Colon, and Kidney;			
RA	Strausberg R.;			
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .			
CC	-----			
DR	EMBL; AF250558; AAF98151.1; -			
DR	EMBL; AF177340; AAG17984.1; -			
DR	EMBL; BC015252; AAH15252.1; -			
DR	EMBL; BC014424; AAH14424.1; -			
DR	InterPro; IPR001832; Claudin.			
DR	InterPro; IPR004031; PMP22_Claudin.			
DR	Pfam; PF00822; PMP22_Claudin; 1.			
DR	PRINTS; PR01077; CLAUDIN.			
DR	PROSITE; PS01346; CLAUDIN; 1.			
KW	Tight junction; Transmembrane.			
FT	TRANSMEM 8 28			POTENTIAL.
FT	TRANSMEM 82 102			POTENTIAL.
FT	TRANSMEM 117 137			POTENTIAL.

```
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA: 24548 MW: 52C6A42DA62B70D CRC64;

Query Match
Best Local Similarity 100.0%; Score 1174; DB 1; Length 230;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
DB 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
QY 61 ITCCDIYSTLLGLPADIOAQAAMVTSSAISLACIISVGMCTVFCQESRAKDRVAVA 120
DB 61 ITCCDIYSTLLGLPADIOAQAAMVTSSAISLACIISVGMCTVFCQESRAKDRVAVA 120
QY 121 GGVFFILGGLGFIPVAMNLHGILRDFYSPLPVDSMKFEIGELALYLGIISSLSFSLAGII 180
DB 121 GGVFFILGGLGFIPVAMNLHGILRDFYSPLPVDSMKFEIGELALYLGIISSLSFSLAGII 180
QY 181 ICFSCSSQNRNSNYDAYOAOPLATRSSPRGQPPKVKSEFNSYSLTGYV 230
DB 181 ICFSCSSQNRNSNYDAYOAOPLATRSSPRGQPPKVKSEFNSYSLTGYV 230

RESULT 2
CUD2_CANFA STANDARD: PRT: 230 AA.
AC 095KM6;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
GN Claudin-2.
OS Cantis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=21206012; PubMed=11309408;
RA Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT Conversion of zonulae occludentes from tight to leaky strand type by
RT introducing claudin-2 into Madin-Darby canine kidney 1 cells.";
RL J. Cell Biol. 153:263-272(2001).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF358907; AAK51433.1; -
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA: 24502 MW: 91B71C1E5DC4BE9 CRC64;

Query Match
Best Local Similarity 93.3%; Score 1119; DB 1; Length 230;
Matches 215; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
```

```
DB 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
QY 61 ITCCDIYSTLLGLPADIOAQAAMVTSSAISLACIISVGMCTVFCQESRAKDRVAVA 120
DB 61 ITCCDIYSTLLGLPADIOAQAAMVTSSAISLACIISVGMCTVFCQESRAKDRVAVA 120
QY 121 GGVFFILGGLGFIPVAMNLHGILRDFYSPLPVDSMKFEIGELALYLGIISSLSFSLAGII 180
DB 121 GGVFFILGGLGFIPVAMNLHGILRDFYSPLPVDSMKFEIGELALYLGIISSLSFSLAGII 180
QY 181 ICFSCSSQNRNSNYDAYOAOPLATRSSPRGQPPKVKSEFNSYSLTGYV 230
DB 181 ICFSCPLQGNRSDYDYDYOAOPLATRGSPRGQPPKAKSEFNSYSLTGYV 230

RESULT 3
CUD2_MOUSE STANDARD: PRT: 230 AA.
AC 088552;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Claudin-2.
GN CLDN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98311639; PubMed=9647647;
RA Furuse M., Fujita K., Hiragi T., Fujimoto K., Tsukita S.;
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
RT tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF072128; AAC27079.1; -
DR MGD: MGI:1276110; Clnd2.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR004031; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA: 24483 MW: 38A7C07A1E0D5D2 CRC64;

Query Match
Best Local Similarity 91.3%; Score 1096; DB 1; Length 230;
Matches 210; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
DB 1 MASLGQLVGYIIGLGLGLTLVAMLLPSWKTSYVGASIVTAAGFSKGLMECATHTSG 60
QY 61 ITCCDIYSTLLGLPADIOAQAAMVTSSAISLACIISVGMCTVFCQESRAKDRVAVA 120
```

```
Db 61 ITQCDIYSTLLGLPADIAQAAMVTSSAMSSSLACIIISVGMRCCTVFCQDSRAKDRVAV 120
QY 121 GGVFFILGGLGFPVAVNHLGILDRFYSPLVPDSMKFGEIGALYGLIISLSFIAGII 180
Db 121 GGVFFILGGLGFPVAVNHLGILDRFYSPLVPDSMKFGEIGALYGLIISLSFIAGVI 180
QY 181 LCFSCSSQRNRSNYDAYQAQPLATRSRPGQPKVKSEFNSYSLTGYV 230
Db 181 LCFSCSQGRTNYDGYQAQPLATRSRPSAQAQPKAKSEFNSYSLTGYV 230

RESULT 4
CLDE_HUMAN STANDARD; PRT; 239 AA.
AC Q95500;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Claudin-14.
GN CLDN14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT DFNB29 ASP-85.
RC TISSUE=Liver;
RX MEDLINE=21097730; PubMed=11163249;
RA Wilcox E.R., Burton Q.L., Naz S., Riazuddin S., Smith T.N.,
RA Plöplis B., Belyantseva I., Ben-Yosef T., Liburd N.A., Morell R.J.,
RA Kachar B., Wu D.K., Griffith A.J., Riazuddin S., Friedman T.B.;
RT "Mutations in the gene encoding tight junction claudin-14 cause
RT autosomal recessive deafness DFNB29."
RL Cell 104:165-172(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Iaudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY. ALSO FOUND IN EAR.
CC -1- DISEASE: DEFECTS IN CLDN14 ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB29).
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AJ132445; CAA10669.1; -
DR EMBL; AF314090; AAG60052.1; -
DR EMBL; BC012126; AAH12126.1; -
DR EMBL; AP001726; BAA95509.1; -
DR MIM; 605608; -
DR InterPro; IPR001832; Claudin.
DR InterPro; IPR004031; PMP22-Claudin.
DR Pfam; PF00822; PMP22-Claudin; 1.
DR PRINTS; PRO1077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane; Disease mutation; Deafness.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT VARIANT 85 85 V->D (IN DFNB29).
FT /FTID=VAR_010738.
SQ SEQUENCE 239 AA; 25699 MW; DD41652F7FD0E09A CRC64;

Query Match 48.6%; Score 571; DB 1; Length 239;
Best Local Similarity 46.8%; Pred No. 2.4e-38;
Matches 111; Conservative 44; Mismatches 70; Indels 12; Gaps 3;

QY 1 MASLGLQLVGYILGLGLGLTLVAMLLPSWKTSYVYGASIVTAVGFSKGLWMECAPHSTG 60
Db 1 MASTAVQLLGLFLSLFLGMVGTLLTILPHRRRTAHVGTNLTAVSYLKGLWMECVHSTG 60
QY 61 ITQCDIYSTLLGLPADIAQAAMVTSSAMSSSLACIIISVGMRCCTVFCQDSRAKDRVAV 120
Db 61 ITQCDIYSTLLGLPADIAQAAMVTSSAMSSSLACIIISVGMRCCTVFCQDSRAKDRVAV 120
QY 121 GGVFFILGGLGFPVAVNHLGILDRFYSPLVPDSMKFGEIGALYGLIISLSFIAGII 180
Db 121 GGVFFILGGLGFPVAVNHLGILDRFYSPLVPDSMKFGEIGALYGLIISLSFIAGII 180
QY 181 LCFSCSSQRNRSNYDAYQAQPLATRSRPGQPKVKSEFNSYSLTGYV 229
Db 181 LCFSCSQGRTNYDGYQAQPLATRSRPSAQAQPKAKSEFNSYSLTGYV 233

RESULT 5
CLDE_MOUSE STANDARD; PRT; 239 AA.
AC Q92053; Q9D284;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Claudin-14.
GN CLDN14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Morita K., Furuse M., Tsukita S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=21097730; PubMed=11163249;
RA Wilcox E.R., Burton Q.L., Naz S., Riazuddin S., Smith T.N.,
RA Plöplis B., Belyantseva I., Ben-Yosef T., Liburd N.A., Morell R.J.,
RA Kachar B., Wu D.K., Griffith A.J., Riazuddin S., Friedman T.B.;
RT "Mutations in the gene encoding tight junction claudin-14 cause
RT autosomal recessive deafness DFNB29."
RL Cell 104:165-172(2001).
RN [3]
RP REVISIONS TO 115; 129; 166 AND 187.
```



```

Db      121 AGVFFLVGGILCLIPVCWSANSVIKDFNPTLSDAQRELQASLFIQCWASGLLLIGGAL 180

Qy      181 LCFSCSSQRNRSNYYDAYOAPLATRRSP 209
        || | | | | | | | | | | | | | | | |
Db      181 LCQCCKPNEGR-----SVYSKISAPRSAP 204

RESULT 7
CLD3_CANFA
ID CLD3_CANFA STANDARD; PRT; 218 AA.
AC O95KM5;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
CLaudin-3.
GN CLDN3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=21206012; PubMed=11309408;
RA Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT *Conversion of zonulae occludentes from tight to leaky strand type by
RT introducing claudin-2 into Madin-Darby canine kidney I cells.";
RL J. Cell Biol. 153:263-272(2001).
CC -1 FUNCTION: COMPONENT OF RIGHT JUNCTION (TJ) STRANDS.
CC -1 SUBCELLULAR LOCATION: Integral membrane protein.
CC -1 SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR ENBL: AF358908; AAK51434.1; -.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 218 AA; 23148 MW; 907104E36F50DA70 CRC64;

Query Match 38.4%; Score 450.5; DB 1; Length 218;
Best Local Similarity 39.6%; Pred. No. 6.9e-29;
Matches 84; Conservative 47; Mismatches 74; Indels 7; Gaps 1;

Qy 3 SLGLQLVGVI TLGLGLGLTLVAMLLPSWKTSYVGVASIVTAVGFSGKLWMECATHSTGIT 62
   |:|:: | | :|| | | | | | | | | | | | | | | | | | | | | |
Db 2 SMGLEINGTSLAVLGWLSTVCCALPMWRVTAFIGSSIITAQTWEGLMNVCVQSTGQM 61

Qy 63 QCDIYSTLLGLPADIAQAQMAYMTSSAIISSLAIIISVGMRCVTFCQESRAKORVAAG 122
   || :|| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 QCKYVDSLALPDQAARALIIVVSILLAAFLGLVALVGAQCTNCVDDTPAKAKITVAG 121

Qy 123 VFETLGLGLGIFPVANLNHGLIRDFYSPLYPSDSMKTEIGEALYLGISSLSIAGLIIC 182
   | :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 VLFLLAALLPLVPVSWSANLTIRDFYNPLVPDQKRMGAGLVYWAAAAAQLLGGALLC 181

Qy 183 FSCSSQRNRSNYYDAYOAPLATRRSPRGQP 214
   || | | | | | | | | | | | | | | | | | | | | | |
Db 182 CSCPPR-----DKKYAPTPIVYSAPRSAGP 206

RESULT 8
CLD3_HUMAN
```

QY 123 VFFILGGLGFIVANNHGLRDFYSPVPSDKFEIGALYGIISLFSLIAGIITLC 182  
DB 122 VFLALALTLVPVSWANSTIRDFYNPVPEAKREMGAGLYVMAAALQLLGGALIC 181  
QY 183 FSCSSORNSRNYDAYAOPLATRRSPRG 212  
DB 182 CSCPP---REKKTATKVVYSAPRST-GPG 207

RESULT 9  
CLD4\_MOUSE STANDARD; PRT: 210 AA.  
AC 035054;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Claudin-4 (Clostridium perfringens enterotoxin receptor) (CPE-receptor) (CPE-R).  
GN CLDN4 OR CPETRI OR CPER.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10990;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97476271; PubMed=9334247;  
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;  
RT "Clostridium perfringens enterotoxin utilizes two structurally related membrane proteins as functional receptors in vivo."  
RL J. Biol. Chem. 272:2652-2658(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99110921; PubMed=9892664;  
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;  
RT "Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands."  
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).  
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AB000713; BAA22985.1; -;  
DR EMBL; AF087822; AAD09757.1; -;  
DR MGJ:1313314; ClDN4.  
DR InterPro: IPR001832; Claudin.  
DR InterPro: IPR004031; PMP22\_Claudin.  
DR Pfam: PF00822; PMP22\_Claudin; 1.  
DR PRINTS: PRO1077; CLAUDIN.  
DR PROSITE: PS01346; CLAUDIN; 1.  
KW Tight junction; Transmembrane.  
FT TRANSMEM 8 28 POTENTIAL.  
FT TRANSMEM 82 102 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
SQ SEQUENCE 210 AA; 22338 MW; 3B6D571EC71D6564 CRC64;

Query Match 38.0%; Score 446; DB 1; Length 210;  
Best Local Similarity 40.2%; Pred. No. 1,5e-28;  
Matches 84; Conservative 46; Mismatches 75; Indels 4; Gaps 1;

QY 1 MASLGLQVGYILGLGLGLTLVAMLLPSKMTSSYVGASIVTAVGFSKGLMECATHTSG 60  
DB 1 MASMGLOVTLGLSLAVLGLGLIISCALPMNRVTAFISGNITVTAQTSMEGLMNNCVQSTG 60

QY 61 ITQCDIYFTLGLPADIQAAQAMWTSSAISLACIISVGNRCTVFCQESRAKDRVAV 120  
DB 61 OMCKMYDSMLALPDQLAARALWVTSIIVGALGMLSVGGKCTCNCHDEDFVAKIMIT 120  
QY 121 GGVFFILGGLGFIVANNHGLRDFYSPVPSDKFEIGALYGIISLFSLIAGIIT 180  
DB 121 AGAVFIVASMLIMVPSWTANHVIRDFYNPVVASGQKREMGASLYVMAASGLILLGGGL 180  
QY 181 LCFSCSSORNSRNYDAYAOPLATRRSP 209  
DB 181 LCCSCPPRSNDKPKYSAKYS----AARVP 205

RESULT 10  
CLD4\_CERAE STANDARD; PRT: 209 AA.  
AC 019005;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Claudin-4 (Clostridium perfringens enterotoxin receptor) (CPE-receptor) (CPE-R).  
GN CLDN4 OR CPETRI OR CPER.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=9534;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97242441; PubMed=9087440;  
RA Katsuhira J., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;  
RT "Molecular cloning and functional characterization of the receptor for Clostridium perfringens enterotoxin."  
RL J. Cell Biol. 136:1239-1247(1997).  
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; D88492; BAA22781.1; -;  
DR InterPro: IPR001832; Claudin.  
DR InterPro: IPR004031; PMP22\_Claudin.  
DR Pfam: PF00822; PMP22\_Claudin; 1.  
DR PRINTS: PRO1077; CLAUDIN.  
DR PROSITE: PS01346; CLAUDIN; 1.  
KW Tight junction; Transmembrane.  
FT TRANSMEM 10 30 POTENTIAL.  
FT TRANSMEM 82 102 POTENTIAL.  
FT TRANSMEM 118 138 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
SQ SEQUENCE 209 AA; 22029 MW; 47ADB3099F95289E CRC64;

Query Match 37.8%; Score 443.5; DB 1; Length 209;  
Best Local Similarity 39.9%; Pred. No. 2,4e-28;  
Matches 83; Conservative 49; Mismatches 71; Indels 5; Gaps 1;

QY 1 MASLGLQVGYILGLGLGLTLVAMLLPSKMTSSYVGASIVTAVGFSKGLMECATHTSG 60  
DB 1 MASMGLOVTLGLAVLAVGLAVMLCCALPMNRVTAFISGNITVTSQITWELMNNCVQSTG 60  
QY 61 ITQCDIYFTLGLPADIQAAQAMWTSSAISLACIISVGNRCTVFCQESRAKDRVAV 120  
DB 61 OMCKMYDSMLALPDQLAARALWVTSIIVGALGMLSVGGKCTCNCHDEDFVAKIMIT 120









Result No.	Score	Query Match	Length	DB	ID	Description
1	1119	95.3	230	6	Q95KM6	Q95km6 canis famill
2	456	38.8	218	13	Q90XR2	Q90xr2 brachydanio
3	450.5	38.4	218	6	Q95KM5	Q95km5 canis famill
4	445	38.0	215	13	Q90XR8	Q90xr8 brachydanio
5	444.5	37.9	209	13	Q90XR9	Q90xr9 brachydanio
6	438	37.3	210	13	Q90XQ8	Q90xq8 brachydanio
7	432.5	36.8	214	13	Q90XQ9	Q90xq9 xenopus lae
8	430	36.6	214	13	Q9dR12	Q9del2 xenopus lae
9	428.5	36.5	214	13	Q98SR2	Q98sr2 gallus gall
10	421	35.9	210	13	Q90XR0	Q90xr0 brachydanio
11	417.5	35.6	210	13	Q90XR4	Q90xr4 brachydanio
12	413.5	35.2	206	13	Q90XS1	Q90xs1 brachydanio
13	396	33.3	211	13	Q90XR6	Q90xr6 brachydanio
14	393.5	33.5	193	11	Q9ET38	Q9et38 mus musculus
15	393	33.5	216	13	Q98SR1	Q98sr1 gallus gall
16	375.5	32.0	200	11	Q91X40	Q91x40 mus musculus

QY 181 LCFSCSSQRNSNYDAYOAOPLATRSPPRGPPKVKSENSYSLTCY 230  
Db 181 LCFSCPLQGNRSYDYDSYAOPLATRSPPRGPPKVKSENSYSLTCY 230

RESULT 2  
Q90XR2 ID Q90XR2 PRELIMINARY; PRT; 218 AA.  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CLAUDIN C.  
GN CLDNC.

OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21417738; PubMed=11517306;  
RA Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359432; AAL01839.1; -;  
SQ SEQUENCE 218 AA; 23425 MW; 82CA23D13306110F CRC64;

Query Match 38.8%; Score 456; DB 13; Length 218;  
Best Local Similarity 38.4%; Pred. No. 3.4e-32;  
Matches 88; Conservative 42; Mismatches 71; Indels 28; Gaps 1;

QY 1 MASLGLOLVGYILGLLGLTGLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHTSG 60  
Db 1 MASFGELGVGLVLSVGLWILNVCCALPMWRVTAFIGNIVTAQVYEGIMWMSCVQSTG 60

QY 61 ITQCDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVGMRCITVFCQESRAKDRVAVA 120  
Db 61 OMCKVYDSMLALPADLOAARLVVAVIIVGLVAFVAVGAKTNCIEEAAKARVWIS 120

QY 121 GGVFFILGGLGFPVAMNHLGILDRFYSPLVPSDMKFEIGEALYLGIISSLSFIAGII 180  
Db 121 SGAFTASVLIQIPVCSAHTVILEFYSPLVPEAKMEIGASLYLGWAASAMLLVGGSI 180

QY 181 LCFSCSSQRNSNYDAYOAOPLATRSPPRGPPKVKSENSYSLTCY 229  
Db 181 LCCSC-----PPKDETTRYPPQSRIAY 201

RESULT 3  
Q95KM5 ID Q95KM5 PRELIMINARY; PRT; 218 AA.  
AC Q95KM5;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE INTEGRAL MEMBRANE PROTEIN CLAUDIN-3.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE=21206012; PubMed=11309408;  
RA Furuse M., Furuse K., Sasaki H., Tsukita S.;  
RT "Conversion of Zonulae Occludentes from Tight to Leaky Strand Type by  
Introducing Claudin-2 into Madin-Darby Canine Kidney I Cells.";  
RL J. Cell Biol. 153:263-272(2001).  
DR EMBL; AF358908; AAK51434.1; -;  
SQ SEQUENCE 218 AA; 23148 MW; 907104E36F50DA70 CRC64;

Query Match 38.4%; Score 450.5; DB 6; Length 218;  
Best Local Similarity 39.6%; Pred. No. 1e-31;  
Matches 84; Conservative 47; Mismatches 74; Indels 7; Gaps 1;

QY 3 SIGLQLVGYILGLLGLTGLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHTSGT 62  
Db 2 SMGLEIAGTSLAVLGLWLTIVCCALPMWRVTAFIGSIITTAQITWGLMNMNVQSTGQM 61

QY 63 QCDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVGMRCITVFCQESRAKDRVAVA 122  
Db 62 CKVYDSMLALPADLOAARLVVAVIIVGLVAFVAVGAKTNCIEEAAKARVWIS 121

QY 123 VFFILGGLGFPVAMNHLGILDRFYSPLVPSDMKFEIGEALYLGIISSLSFIAGII 182  
Db 123 VFLFLAALLTLVPVGSANTIRDPNPLVPAQOREMGAGLYGVWAAAALLOGLGALLC 181

QY 183 FSCSSQRNSNYDAYOAOPLATRSPPRGQP 214  
Db 182 CSCPPR-----DKYAPTKEYSAPRSAGP 206

RESULT 4  
Q90XR8 ID Q90XR8 PRELIMINARY; PRT; 215 AA.  
AC Q90XR8;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CLAUDIN B.  
GN CLDNB.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21417738; PubMed=11517306;  
RA Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359426; AAL01833.1; -;  
SQ SEQUENCE 215 AA; 22656 MW; F3D944F527751145 CRC64;

Query Match 38.0%; Score 446; DB 13; Length 215;  
Best Local Similarity 39.7%; Pred. No. 2.5e-31;  
Matches 83; Conservative 45; Mismatches 81; Indels 0; Gaps 0;

QY 1 MASLGLOLVGYILGLLGLTGLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHTSG 60  
Db 1 MASTGLQMLGIALAIFGWIIVLCAALPMWRVTAFIGNIVTSQTSWEGIMWMSCVQSTG 60

QY 61 ITQCDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVGMRCITVFCQESRAKDRVAVA 120  
Db 61 OMCKVYDSMLALSDIOAARALTIVISIVGMWIMLWAGKTCNCIEEESKARVGT 120

QY 121 GGVFFILGGLGFPVAMNHLGILDRFYSPLVPSDMKFEIGEALYLGIISSLSFIAGII 180  
Db 121 AGVFIISGVLCVPCVTANAIQDFYNPLVPAQOREMGAGLYGVWAAAALLOGLGALLC 180

QY 181 LCFSCSSQRNSNYDAYOAOPLATRSPP 209  
Db 181 LCCHCPEKSDSKYTAKNATPRSEASAP 209

RESULT 5  
Q90XR9 ID Q90XR9 PRELIMINARY; PRT; 209 AA.  
AC Q90XR9;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)

```

DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE CLAUDIN E.
GN CLDNE.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP
RP MEDLINE=21417738; PubMed=11517306;
RA Kollmer R., Nakamura S.K., Kappeler J.A., Hudspeith A.J.;
RT "Expression and phylogeny of claudins in vertebrate primordia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR EMBL; AF359425; AAL01832.1;
SQ SEQUENCE 209 AA; 22593 MW; A66FA8D852E16CB2 CRC64;

Query Match 37.9%; Score 444.5; DB 13; Length 209;
Best Local Similarity 40.4%; Pred. No. 3.3e-31;
Matches 84; Conservative 49; Mismatches 72; Indels 3; Gaps 1;

OY 1 MASLQLQVYITLGLGLGLTLVAMLLPSKRTSSVGVASIVAVGFSKGLMEKATHSG 60
Db 1 MYSKREILGMCALITGFLGAILICALPMPKVTATGATVIAQTWESLMNCVQMSHG 60
OY 61 ITQCDIVSTLGLPADIQAAQAMMTSSAIISSACIISVGVGRCTVFCESRAKDVAVA 120
Db 61 QMCKRTVSLALPDQDAARLVAIVATVSPFALLIGAGCKCNFVREDAKKAVSIA 120
OY 121 GGVFFILGLGLFTIVAMNLHGLIDFVSPVPSDKMEFEGALYLGIISSLFSLAGII 180
Db 121 SGVIFITGVILVLPVCSANTIIIDFNPILTDQOREMGASLYIGWVAALLIGGI 180
OY 181 LCFSSSQGRNSNTYDAYQAOPLARRSS 208
Db 181 LCSCPPPKDER--YNMKYSOPRSTRATS 205

RESULT 6
O90X08 PRELIMINARY: PRT: 210 AA.
AC O90X08:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLAUDIN 19.
GN CLDNI9.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP
RP MEDLINE=21417738; PubMed=11517306;
RA Kollmer R., Nakamura S.K., Kappeler J.A., Hudspeith A.J.;
RT "Expression and phylogeny of claudins in vertebrate primordia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).
DR EMBL; AF359436; AAL01843.1;
SQ SEQUENCE 210 AA; 22253 MW; FEFB5ICE7A98D47 CRC64;

Query Match 37.3%; Score 438; DB 13; Length 210;
Best Local Similarity 41.3%; Pred. No. 1.2e-30;
Matches 88; Conservative 37; Mismatches 80; Indels 8; Gaps 2;

OY 1 MASLQLQVYITLGLGLTLVAMLLPSKRTSSVGVASIVAVGFSKGLMEKATHSG 60
Db 1 MAHAGLQWLGICLGLGLGLGLTIAAMBMKSSAGDNIITTAQOYELGMSCVSOSTG 60
OY 61 ITQCDIVSTLGLPADIQAAQAMMTSSAIISSACIISVGVGRCTVFCOES--RAKDVAV 119

```

Query Match	36.8%; Score 432.5; DB 13; Length 214;
Best Local Similarity	38.6%; Pred. No. 3.7e-30;
Matches	80; Conservative 50; Mismatches 76; Indels 1; Gaps 1;
Db 61	QLOCCKYDSLKLPGELIQGARGMLTGLFLGLSLTVSVFGMKCTTGCLSEAPQVNSKVAL 120
Qy 120	AGVFFIITGLGLGFIIVANNLHGLIRDFYSPVLPDSMKREIEALYLGITSSLSFLIAGI 179
Db 121	AGGVFLITIGGLFALLATSMYGEKINQKFFDPPTPINAREPFGKALYVGMSALSITIGS 180
Qy 180	ILCFSCSSSORNSNRYDAYOAPLATRSSPRG 212
Db 181	ILCCICGSEASEKPSYP-----PARAAGRPG 206
RESULT 7	
ID 090X09	PRELIMINARY; PRT; 214 AA.
AC 090X09;	
DT 01-DEC-2001 (TREMBLrel. 19, Created)	
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE CLAUDIN A.	
GN CLDNA.	
OS Xenopus laevis (African clawed frog).	
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;	
OC Xenopodinae; Xenopus.	
NCBI_TaxID=8355;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=21417738; PubMed=11517306;	
RA Kolimar R., Nakamura S.K., Kappeler J.A., Hudspeth A.J.;	
RT "Expression and phylogeny of claudins in vertebrate primordia."	
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).	
DR EMBL; AF359435; AAL01842.1; "	
SQ SEQUENCE 214 AA; 22952 MW; 8EB6F0EB2B7257D CRC64;	

RC TISSUE=DORSAL LIP;  
RX MEDLINE=21098758; PubMed=11161574;  
RA Brizuela B.J., Wessely O., De Robertis E.M.;  
RT "Overexpression of the xenopus Tight-Junction Protein Claudin Causes  
Randomization of the Left-Right Body Axis.";  
RL Dev. Biol. 230:217-229(2001).  
DR EMBL; AF224712; AAG44257.1; -  
DR InterPro; IPR001832; Claudin.  
DR InterPro; IPR004031; PMP22\_Claudin.  
DR Pfam; PF00822; PMP22\_Claudin; 1.  
DR PRINTS; PR01077; CLAUDIN.  
DR PROSITE; PS01346; CLAUDIN; UNKNOWN\_1.  
SQ SEQUENCE 214 AA; 23151 MW; 3F0D65E3ADE311E6 CRC64;

Query Match 36.6%; Score 430; DB 13; Length 214;  
Best Local Similarity 36.5%; Pred. No. 6.2e-30;  
Matches 84; Conservative 52; Mismatches 78; Indels 16; Gaps 3;

QY 1 MASLGQLVGYIIGLLGLTLVAMLLPSWKTSYVGASIVTAVGFSKGLWMECATHTSGT 60  
Db 1 MASTGLQLGMAMSIIGWGSIIISCALPMRVTAFIGNNIVVAQIIWEGLWMNCIVQSTG 60

QY 61 ITQCDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVVGMRCTVFCQESRAKDRVAV 120  
Db 61 OMCKYVDSMLAQLDQAAALWVAIVLAVGLMVAIVGAQCRCVDETTKAKITIVSG 120

QY 121 GGVFILGGLGTFIPVAVNHLGILRDFYSPVLPDSMKFEIGALYLGIIISLFLSIAGII 180  
Db 121 SGIVFLVAGILMLIPVCWSANSIIRDFYNPLVPEAQRELGAAIYIRWASSALLLGGSL 180

QY 181 LCFSCSSQRNRSNYDAYOAPLATRSSRPGOPPKVKSEFNSYSLRGYV 230  
Db 181 LCCSC-PKKNADYPARYTA-----PSCPPR-----SDYTSKNYV 214

RESULT 9  
Q98SR2 ID Q98SR2 PRELIMINARY; PRT; 214 AA.  
AC Q98SR2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE CLAUDIN-3.  
GN Gallus gallus (Chicken).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Reardon E., Kojima S., Rizzolo L.J.;  
RT "Sequence of chick claudin-3 cDNA.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF34677; AAK20876.1; -  
DR InterPro; IPR001832; Claudin.  
DR InterPro; IPR004031; PMP22\_Claudin.  
DR Pfam; PF00822; PMP22\_Claudin; 1.  
DR PRINTS; PR01077; CLAUDIN.  
DR PROSITE; PS01346; CLAUDIN; UNKNOWN\_1.  
SQ SEQUENCE 214 AA; 23090 MW; A2540116CBD53978 CRC64;

Query Match 36.5%; Score 428.5; DB 13; Length 214;  
Best Local Similarity 36.4%; Pred. No. 8.4e-30;  
Matches 83; Conservative 50; Mismatches 80; Indels 15; Gaps 2;

QY 3 SLGLQLVGYIIGLLGLTLVAMLLPSWKTSYVGASIVTAVGFSKGLWMECATHTSGIT 62  
Db 3 SMGLEIGGVALSVIGWLCISICCALPMRVTAFIGNNIVVAQIIWEGLWMNCIVVOSTGOM 61

QY 63 QCIDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVVGMRCTVFCQESRAKDRVAVAG 122

Db 62 QCKYVDSMLALPQDLQAARALLWVAIVLAVGLMVAIVGAQCRCVDETTKAKITIVSG 121

QY 123 VFFILGGLGTFIPVAVNHLGILRDFYSPVLPDSMKFEIGALYLGIIISLFLSIAGIILC 182  
Db 123 VFFILGGLGTFIPVAVNHLGILRDFYSPVLPDSMKFEIGALYLGIIISLFLSIAGIILC 182

QY 183 FSCSSQRNRSNYDAYOAPLATRSSRPGOPPKVKSEFNSYSLRGYV 230  
Db 183 CSCPPK-----DERYAPSKVAYSAPR-----SAVTSYDKRNYV 214

RESULT 10  
Q90XR0 ID Q90XR0 PRELIMINARY; PRT; 214 AA.  
AC Q90XR0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CLAUDIN H.  
GN CLDNH.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=21417738; PubMed=11517306;  
RA Kollmar R., Nakamura S.K., Kappler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359434; AAL01841.1; -  
SQ SEQUENCE 214 AA; 22788 MW; 38163AB7E50BF058 CRC64;

Query Match 35.9%; Score 421; DB 13; Length 214;  
Best Local Similarity 38.5%; Pred. No. 3.8e-29;  
Matches 80; Conservative 49; Mismatches 65; Indels 14; Gaps 1;

QY 3 SLGLQLVGYIIGLLGLTLVAMLLPSWKTSYVGASIVTAVGFSKGLWMECATHTSGIT 62  
Db 3 SMGLEIGGVALSVIGWLCISICCALPMRVTAFIGNNIVVAQIIWEGLWMNCIVVOSTGOM 61

QY 63 QCIDIYSTLLGLPADIOAAQAMMTSSAIISSLACIIISVVGMRCTVFCQESRAKDRVAVAG 122  
Db 62 QCKYVDSMLALPQDLQAARALLWVAIVLAVGLMVAIVGAQCRCVDETTKAKITIVSG 121

QY 123 VFFILGGLGTFIPVAVNHLGILRDFYSPVLPDSMKFEIGALYLGIIISLFLSIAGIILC 182  
Db 123 IMFIAGIILDLPSAWANQIIRDFYNPLPGAQOQRELGAIIYIGFAAALLIIGGAMLC 181

QY 183 FSC-----SSQRNRSNYD 196  
Db 182 CTCPPKPKKYPARMGYSAPRSASAGYD 209

RESULT 11  
Q90XR4 ID Q90XR4 PRELIMINARY; PRT; 210 AA.  
AC Q90XR4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CLAUDIN J.  
GN CLDNJ.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21417738; Pubmed=11517306;  
RA Kollmar R., Nakamura S.K., Kappeler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359430; AAL01837.1; -  
SQ SEQUENCE 210 AA; 22895 MW; F9A6964463E55775 CRC64;

Query Match 35.68; Score 417.5; DB 13; Length 210;  
Best Local Similarity 36.14; Pred. No. 7.4e-29;  
Matches 82; Conservative 47; Mismatches 81; Indels 17; Gaps 2;

QY 4 LGQIVYIIIGLGLLTIVAMLLPSKTSYVGSITANGFSGKLMMECATHTSGITQ 63  
DB 1 MALQVILITISMIGFAITIIICALPMKVTFTIGTIVAAOVFEGIMMTCYERIGQM 60  
QY 64 CDISTYLLGLPADIOAQAAMVTSATSLACITSVGMRCVFCQESRAKDRVAVAGV 123  
DB 61 CKLYDALDDLPLOASRGILVTMALASLAFILFIDGADCTNCLSNFRAKRIVVSGI 120  
QY 124 FTLLGLLGFIPVAMNLHGILRDFYSPVPSMKFEIGELALYGIISLFLIACIILCF 183  
DB 121 TFMISGLTIVVPSWTFADSIIRDFHNPVHVALKREMGALYVGMVLTAGFLFVGALLICT 180  
QY 184 SCSSQRNRSNYYDAYOAPLATRSPRGQPPKXSEFNSTSLGCV 230  
DB 181 SCPPERD--NYLPRYTLTKSGTHS-----GAVKNYV 210

## RESULT 12

ID Q90XS1 PRELIMINARY; PRT; 206 AA.  
AC Q90XS1;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
GN CLAUDIN A.  
OS Brachydanio rerio (zebrafish) (zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21417738; Pubmed=11517306;  
RA Kollmar R., Nakamura S.K., Kappeler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359423; AAL01830.1; -  
SQ SEQUENCE 206 AA; 21694 MW; 69CF7A8554A2EFCF CRC64;

Query Match 35.28; Score 413.5; DB 13; Length 206;  
Best Local Similarity 37.4%; Pred. No. 1.6e-28;  
Matches 77; Conservative 46; Mismatches 80; Indels 3; Gaps 1;

QY 1 MASLGIQVYIIIGLGLLTIVAMLLPSKTSYVGSITANGFSGKLMMECATHTSG 60  
DB 1 MVSAGIQVGVIALAVIGMIGVYICLLPMQVITAFICQNTIYAQVYTWGIMMSCVYSTG 60  
QY 61 ITQCDIYSTLLGLPADIOAQAAMVTSATSLACITSVGMRCVFCQESRAKDRVAV 120  
DB 61 QMCKYVDSMLALSSDQARALCITISLVGVGILLAAAGKCTTICXERAKRKYCVI 120  
QY 121 GGVFTLLGLLGFIPVAMNLHGILRDFYSPVPSMKFEIGELALYGIISLFLIACI 180  
DB 121 SGALFIAGVYICLIPVCMWANNIRDFYNPMTNSAQKRELQASLFIGASALLIIGSL 180  
QY 181 LCFSQSSQRNRSNYYDAYOAPLATR 206  
DB 181 LCANCPQ--DOYKATYTRARSGTK 203

RESULT 13  
ID Q90XR6 PRELIMINARY; PRT; 211 AA.  
AC Q90XR6;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DE CLAUDIN I.  
GN CLDN1.  
OS Brachydanio rerio (zebrafish) (zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21417738; Pubmed=11517306;  
RA Kollmar R., Nakamura S.K., Kappeler J.A., Hudspeth A.J.;  
RT "Expression and phylogeny of claudins in vertebrate primordia."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10196-10201(2001).  
DR EMBL; AF359428; AAL01835.1; -  
SQ SEQUENCE 211 AA; 22174 MW; 60F9E453891B5639 CRC64;

Query Match 33.7%; Score 396; DB 13; Length 211;  
Best Local Similarity 38.3%; Pred. No. 5.6e-27;  
Matches 79; Conservative 44; Mismatches 79; Indels 4; Gaps 2;

QY 1 MASLGIQVYIIIGLGLLTIVAMLLPSKTSYVGSITANGFSGKLMMECATHTSG 60  
DB 1 MGSAGQIVCVALLIGLGLLTIVVTAIPQWKTSAFIGQNTITFQVSEEGIMQCVYSTG 60  
QY 61 ITQCDIYSTLLGLPADIOAQAAMVTSATSLACITSVGMRCVFCQESRAKDRVAV 120  
DB 61 QQQCSTYDLSLLISDQARARFTITSCMLSTVSLILCGADFTCIENEDVKPVVTV 120  
QY 121 GGVFTLLGLLGFIPVAMNLHGILRDFYSPVPSMKFEIGELALYGIISLFLIACI 179  
DB 121 SAILGILAGLVIIIPVSMANNVVRFPNPMVPEAKRELGPOTYIGMASVILLIAGL 180  
QY 180 ILCFS---CSSQRNRSNYYDAYOAP 202  
DB 181 LCFSRPRSSGSSGAAYSNSASAP 206

## RESULT 14

ID Q9ET38 PRELIMINARY; PRT; 193 AA.  
AC Q9ET38;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DE 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
GN CLAUDIN-19 (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR;  
RA Kluethi Y., Morita K., Furuse M., Tsukita S.;  
RT Submitted (MAR-2000) to the EMBL/Genbank/DBD databases.  
DR EMBL; AF249888; AAF98323.1; -  
DR InterPro; IPR001832; Claudin.  
DR InterPro; IPR004031; PMP22.Claudin.  
DR Pfam; PF00882; PMP22.Claudin; 1.  
DR PRINTS; PR01077; CLAUDIN.  
DR PROSITE; PS01346; CLAUDIN; UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 193  
SQ SEQUENCE 193 AA; 20299 MW; 2F2D82DB5FCF0D7F CRC64;

Job time: 938 sec

Query Match 33.5%; Score 393.5; DB 11; Length 193;  
Best Local Similarity 39.4%; Pred. No. 8.4e-27;  
Matches 76; Conservative 41; Mismatches 75; Indels 1; Gaps 1;

QY 11 YILGLLGLLGLTGLVAMLLPSWKTSYVGASIVTAVGFSKGLWMECATHSTGITQCDIYSTL 70  
Db 1 YFLALGGWGIITASTALPQWKQSYAGDAIITAVGLYEGWLWMCASOSTGOVKLYDSL 60

QY 71 LGLPADIAQAAMMTSSAISLACIISVVGMRCT-VFCQESRAKDRVAVAGGVFFILGG 129  
Db 61 LALDGHQISARALMVAVLIGFVAMVLSVVGMRCTRVGDSNPTAKSRVAISGGALFLLAG 120

QY 130 LLGFIPVANNHGLIRDFFSPLVPDSMKFEIGEALYLGIISSLSFLSIAGIILCFSCSSQR 189  
Db 121 LCTLTAVSWYATLVTOEFFNPSTPVNARYEFGPALFVGWASAGLAMLGGSFLLCCTCPEPE 180

QY 190 NRSNYDAYOAOQ 202  
Db 181 RANSIPOPYNSGP 193

RESULT 15  
Q98SR1 PRELIMINARY; PRT; 216 AA.

ID Q98SR1  
AC Q98SR1;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE CLAUDIN-5.  
GN CLDN5.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Reardon E., Kojima S., Rizzolo L.J.;  
RT "Sequence of chick claudin-5 cDNA";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF334678; AAK20877.1; -  
DR InterPro; IPR001832; Claudin.  
DR pfam; PF004031; PMP22\_Claudin.  
DR Pfam; PF00822; PMP22\_Claudin; 1.  
DR PRINTS; PR01077; CLAUDIN.  
DR PROSITE; PS01346; CLAUDIN; UNKNOWN\_1.  
SQ SEQUENCE 216 AA; 23180 MW; 5037A882DFF5433A CRC64;

Query Match 33.5%; Score 393; DB 13; Length 216;  
Best Local Similarity 36.2%; Pred. No. 1e-26;  
Matches 75; Conservative 48; Mismatches 84; Indels 0; Gaps 0;

QY 1 MASLGLQVGYILGLLGLTGLVAMLLPSWKTSYVGASIVTAVGFSKGLWMECATHSTG 60  
Db 1 MASAAVEILGLGLGILGWGVILACGLPMQVSAFIDNIVVAQTIVEGLWMCVQSTG 60

QY 61 ITQCDIYSTLGLPADIAQAAMMTSSAISLACIISVVGMRCTVFCQESRAKDRVAVA 120  
Db 61 QMOCKVYDSILALRPEVQAGRALTVIVALGLVALMVTYVGAOCTNCRPGKMSRVIA 120

QY 121 GGVEFFILGLGLFIPVANNHGLIRDFFSPLVPDSMKFEIGEALYLGIISSLSFLSIAGII 180  
Db 121 GGTIYILCGVLVLPICWFANIVISDFPSPQKREIGAALYIGWAATALLFGGCL 180

QY 181 LCFSCSSQRNRSNYDAYOAOPLATRS 207  
Db 181 ICCSCCLQRDETSPVKYSAPRRPTSS 207

Search completed: June 2, 2002, 20:15:55

